162/13.8.

Vol. 62, No. 1 January, 1955

Psychological Review

THEODORE M. NEWCOMB, Editor
UNIVERSITY OF MICHIGAN
Lorraine Bouthilet, Managing Editor

478z40

CONTENTS

Breadth of Learning as a Function of Drive Level and Mechanization.....JEROME S. BRUNER, JEAN MATTER, AND MIRIAM LEWIN PAPANEK 1
A Theory of Discrimination Learning................FRANK RESTLE 11
A Stimulus-Response Analysis of the Interaction of Cue-Producing and Instrumental Responses...............................ALBERT E. GOSS 20
Perceptual Learning: Differentiation or Enrichment....................................JAMES J. GIBSON AND ELEANOR J. GIBSON 32
The Principle of Congruity in the Prediction of Attitude Change.....................CHARLES E. OSGOOD AND PERCY H. TANNENBAUM 42
A Model of Short- and Long-Run Mechanisms Involved in Pressures Toward Uniformity in Groups...................HERBERT A. SIMON AND HAROLD GUETZKOW 56
The Concept of Efficiency in Psychological Health and in Psychopathology..............JULIUS WISHNER 69

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

OF EDUC. & PSYCH.
LIBRARY

mk
/917

CONSULTING EDITORS

Solomon E. Asch
Robert R. Blake
Stuart W. Cook
Clyde H. Coombs
Leon Festinger
W. R. Garner
James J. Gibson
D. O. Hebb
Harry Helson
E. R. Hilgard
Carl I. Hovland
E. Lowell Kelly
David Krech
Robert W. Leeper

Robert B. MacLeod
David C. McClelland
George A. Miller
Gardner Murphy
Oscar Oeser
Carl Pfaffmann
Carroll C. Pratt
David Shakow
Richard L. Solomon
Eliot Stellar
S. S. Stevens
Eric Trist
Edward L. Walker
Robert W. White

13-8-70
8162

The *Psychological Review* is devoted to theoretical articles of significance to any area of psychology. Except for occasional articles solicited by the Editor, manuscripts exceeding twelve printed pages (about 7,500 words) are not accepted. Ordinarily manuscripts which consist primarily of original reports of research should be submitted to other journals.

Because of the large number of manuscripts submitted, there is an inevitable publication lag of several months. Authors may avoid this delay if they are prepared to pay the costs of publishing their own articles; the appearance of articles by other contributors is not thereby delayed.

Tables, footnotes, and references should appear on separate pages; all of these, as well as the text, should be typed double-spaced throughout, in all manuscripts submitted. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail. Manuscripts should be addressed to the Editor, Dr. Theodore M. Newcomb, Doctoral Program in Social Psychology, University of Michigan, Ann Arbor, Michigan.

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STREETS, LANCASTER, PA.
1333 SIXTEENTH ST. N. W., WASHINGTON 6, D. C.

$6.50 volume $1.50 issue

Entered as second-class matter July 13, 1897, at the post-office at Lancaster, Pa., under Act of Congress of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized Jan. 8, 1948

Send all communications, including address changes, to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must arrive by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Copyright 1955 by the American Psychological Association, Inc.

Vol. 62, No. 3 May, 1955

Psychological Review

THEODORE M. NEWCOMB, Editor
UNIVERSITY OF MICHIGAN
Lorraine Bouthilet, Managing Editor

CONTENTS

Statistical Theory of Spontaneous Recovery and Regression.....W. K. ESTES 145
The Constancy of the *I–V:* A Critique of Intervening Variables.....JOHN P. SEWARD 155
Pleasure-Seeking Behavior and the Drive-Reduction Hypothesis.....JUDSON S. BROWN 169
The Functions of Schemata in Perceiving.....M. D. VERNON 180

SYMPOSIUM ON THE PROBABILITY APPROACH IN PSYCHOLOGY

Representative Design and Probabilistic Theory in a Functional Psychology.....EGON BRUNSWIK 193
The Probability Approach and Nomothetic Theory.....LEO POSTMAN 218
Discussion of Probabilistic Functionalism.....ERNEST R. HILGARD 226
Discussion: Theory and Reductionism.....DAVID KRECH 229
Functionalism, Psychological Theory, and the Uniting Sciences: Some Discussion Remarks.....HERBERT FEIGL 232
In Defense of Probabilistic Functionalism: A Reply.....EGON BRUNSWIK 236

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

mle
917

CONSULTING EDITORS

Solomon E. Asch
Robert R. Blake
Stuart W. Cook
Clyde H. Coombs
Leon Festinger
W. R. Garner
James J. Gibson
D. O. Hebb
Harry Helson
E. R. Hilgard
Carl I. Hovland
E. Lowell Kelly
David Krech
Robert W. Leeper
Robert B. MacLeod
David C. McClelland
George A. Miller
Gardner Murphy
Oscar Oeser
Carl Pfaffmann
Carroll C. Pratt
David Shakow
Richard L. Solomon
Eliot Stellar
S. S. Stevens
Eric Trist
Edward L. Walker
Robert W. White

The *Psychological Review* is devoted to theoretical articles of significance to any area of psychology. Except for occasional articles solicited by the Editor, manuscripts exceeding twelve printed pages (about 7,500 words) are not accepted. Ordinarily manuscripts which consist primarily of original reports of research should be submitted to other journals.

Because of the large number of manuscripts submitted, there is an inevitable publication lag of several months. Authors may avoid this delay if they are prepared to pay the costs of publishing their own articles; the appearance of articles by other contributors is not thereby delayed.

Tables, footnotes, and references should appear on separate pages; all of these, as well as the text, should be typed double-spaced throughout, in all manuscripts submitted. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail. Manuscripts should be addressed to the Editor, Dr. Theodore M. Newcomb, Doctoral Program in Social Psychology, University of Michigan, Ann Arbor, Michigan.

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STREETS, LANCASTER, PA.
1333 SIXTEENTH ST. N. W., WASHINGTON 6, D. C.
$6.50 volume
$1.50 issue

Entered as second-class matter July 13, 1897, at the post-office at Lancaster, Pa., under Act of Congress of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized Jan. 8, 1948

Send all communications, including address changes, to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must arrive by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Copyright 1955 by the American Psychological Association, Inc.

Vol. 62, No. 4 July, 1955

Psychological Review

THEODORE M. NEWCOMB, Editor
UNIVERSITY OF MICHIGAN
Lorraine Bouthilet, Managing Editor

CONTENTS

Drives and the C.N.S. (Conceptual Nervous System) .... D. O. HEBB 243

Probabilistic Functioning and the Clinical Method .... KENNETH R. HAMMOND 255

Feedback Theory and the Reflex Arc Concept ...... CHARLES W. SLACK 263

Interpersonal Behavior as Influenced by Accuracy of Social Perception .... IVAN D. STEINER 268

Thinking: From a Behavioristic Point of View ..... IRVING MALTZMAN 275

Visual Figure Discrimination and the Mediation of Equivalence Responses .... A. L. TOWE 287

Elicitation Theory: I. An Analysis of Two Typical Learning Situations .... M. RAY DENNY AND HARVEY M. ADELMAN 290

The Psychology of Mental Content Reconsidered .... DAVID C. McCLELLAND 297

On Making Predictions from Hull's Theory .... JOHN W. COTTON 303

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

CONSULTING EDITORS

Solomon E. Asch
Robert R. Blake
Stuart W. Cook
Clyde H. Coombs
Leon Festinger
W. R. Garner
James J. Gibson
D. O. Hebb
Harry Helson
E. R. Hilgard
Carl I. Hovland
E. Lowell Kelly
David Krech
Robert W. Leeper
Robert B. MacLeod
David C. McClelland
George A. Miller
Gardner Murphy
Oscar Oeser
Carl Pfaffmann
Carroll C. Pratt
David Shakow
Richard L. Solomon
Eliot Stellar
S. S. Stevens
Eric Trist
Edward L. Walker
Robert W. White

The *Psychological Review* is devoted to theoretical articles of significance to any area of psychology. Except for occasional articles solicited by the Editor, manuscripts exceeding twelve printed pages (about 7,500 words) are not accepted. Ordinarily manuscripts which consist primarily of original reports of research should be submitted to other journals.

Because of the large number of manuscripts submitted, there is an inevitable publication lag of several months. Authors may avoid this delay if they are prepared to pay the costs of publishing their own articles; the appearance of articles by other contributors is not thereby delayed.

Tables, footnotes, and references should appear on separate pages; all of these, as well as the text, should be typed double-spaced throughout, in all manuscripts submitted. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail. Manuscripts should be addressed to the Editor, Dr. Theodore M. Newcomb, Doctoral Program in Social Psychology, University of Michigan, Ann Arbor, Michigan.

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STREETS, LANCASTER, PA.
1333 SIXTEENTH ST. N. W., WASHINGTON 6, D. C.
$6.50 volume
$1.50 issue

Entered as second-class matter July 13, 1897, at the post-office at Lancaster, Pa., under Act of Congress of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized Jan. 8, 1948

Send all communications, including address changes, to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must arrive by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Copyright 1955 by the American Psychological Association, Inc.

Vol. 62, No. 5 September, 1955

Psychological Review

THEODORE M. NEWCOMB, Editor
UNIVERSITY OF MICHIGAN
Lorraine Bouthilet, Managing Editor

CONTENTS

Principles of Performance........................EDWARD C. TOLMAN 315
A Note on Second-Order Conditioning—and Secondary Reinforcement....................GREGORY RAZRAN 327
Psychology from a Biologist's Point of View..............................C. JUDSON HERRICK 333
The Stimulus Conditions Which Follow Learned Responses....................CHARLES C. PERKINS, JR. 341
Consciousness, Behavior, and Personality..................................JOSEPH NUTTIN 349
The Difference Between *Q* and *R*.......................JACK BLOCK 356
Some Problems in the Theory of Vigilance..............................JAMES DEESE 359
Statistical Theory of Distributional Phenomena in Learning......................W. K. ESTES 369
Relationships Between Shame and Guilt in the Socializing Process......................DAVID P. AUSUBEL 378
Notes Toward a Motor Theory of Visual Egocentric Localization......................JAN H. BRUELL AND GEORGE W. ALBEE 391

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

EDUCATIONAL & PSYCHOLOGICAL RESEARCH
DAVID HARE TRAINING COLLEGE
CALCUTTA
19/11/5

CONSULTING EDITORS

Solomon E. Asch
Robert R. Blake
Stuart W. Cook
Clyde H. Coombs
Leon Festinger
W. R. Garner
James J. Gibson
D. O. Hebb
Harry Helson
E. R. Hilgard
Carl I. Hovland
E. Lowell Kelly
David Krech
Robert W. Leeper
Robert B. MacLeod
David C. McClelland
George A. Miller
Gardner Murphy
Oscar Oeser
Carl Pfaffmann
Carroll C. Pratt
David Shakow
Richard L. Solomon
Eliot Stellar
S. S. Stevens
Eric Trist
Edward L. Walker
Robert W. White

The *Psychological Review* is devoted to theoretical articles of significance to any area of psychology. Except for occasional articles solicited by the Editor, manuscripts exceeding twelve printed pages (about 7,500 words) are not accepted. Ordinarily manuscripts which consist primarily of original reports of research should be submitted to other journals.

Because of the large number of manuscripts submitted, there is an inevitable publication lag of several months. Authors may avoid this delay if they are prepared to pay the costs of publishing their own articles; the appearance of articles by other contributors is not thereby delayed.

Tables, footnotes, and references should appear on separate pages; all of these, as well as the text, should be typed double-spaced throughout, in all manuscripts submitted. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail. Manuscripts should be addressed to the Editor, Dr. Theodore M. Newcomb, Doctoral Program in Social Psychology, University of Michigan, Ann Arbor, Michigan.

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STREETS, LANCASTER, PA.
1333 SIXTEENTH ST. N. W., WASHINGTON 6, D. C.

$6.50 volume $1.50 issue

Entered as second-class matter July 13, 1897, at the post-office at Lancaster, Pa., under Act of Congress of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized Jan. 8, 1948

Send all communications, including address changes, to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must arrive by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Copyright © 1955 by the American Psychological Association, Inc.

VOL. 62, NO. 1 JANUARY, 1955

THE PSYCHOLOGICAL REVIEW

478 240.

BREADTH OF LEARNING AS A FUNCTION OF DRIVE LEVEL AND MECHANIZATION [1]

JEROME S. BRUNER, JEAN MATTER, AND MIRIAM LEWIN PAPANEK

Harvard University

Suppose a group of rats is trained to respond to the black doors of a four-unit discrimination apparatus, each unit comprising a black and a white door side by side. (See Fig. 1.) And suppose further that these correct doors, during learning, are arranged from trial to trial in a random pattern with respect to whether they are on the left or the right. Under these circumstances, the animals would be learning to respond to the cue of door brightness.

Now suppose, after the animals have learned the black-white discrimination to a criterion, the correct black doors are arranged for a series of trials in a pattern of single alternation. That is, the correct black door is always in the sequence of left-right-left-right. We are interested in finding out how much the animals learn about single alternation during this period when the original black cue is present along with the single-alternation pattern. To test for this learning of the single-alternation pattern, we then remove the black-white discriminanda, inserting gray doors instead, and see how many trials are required for the rats to master the single-alternation pattern to criterion. A control group is required, of course: animals that continue with randomly positioned black-white training for a number of trials comparable to the series of trials during which experimental animals receive the training with the double cue of black and single alternation.

It would appear at first that such an experiment addresses itself directly to testing predictions that might be derived from the continuity and discontinuity hypotheses about discrimination learning. The continuity hypothesis might hold, to put the matter in somewhat simplified terms, that insofar as reinforcement follows the single-alternation responses during the critical period when black-white cues are still present, then the animal will learn this alternation response even though his original training was to respond to the black doors. Put in more refined terms, reinforcement will lead to the learning of the single-alternation response, provided a sufficient number of reinforced trials has been given. The discontinuity position might hold that insofar as the organism is responding in terms of a black-going hypothesis, it does not learn anything about single alternation during the critical period when the relevant black cue is still present and a black-going hypothesis is adequate for reaching the goal.

The difference between these two classical positions can, we will hold, be reduced to a question about the *range of cues* to which the animal is responding in the course of discrimination

---

[1] We are indebted to Professor E. C. Tolman for many stimulating suggestions, and to Professors C. F. Mosteller, R. R. Bush, L. Postman, and D. Krech for valuable advice and criticism.

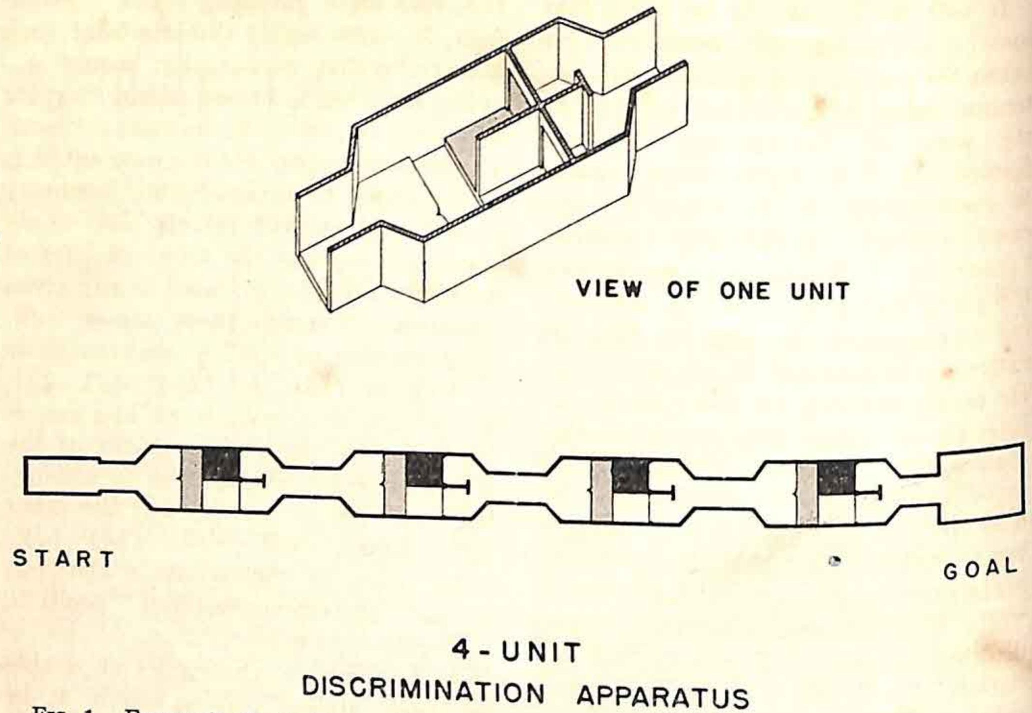


FIG. 1. Four-unit discrimination apparatus (after Greenbaum). Length of each unit is 28 in., maximum width of each unit 9 in., minimum width 4 in. The height of the alleys is 4 in. The entire apparatus is covered by a Plexiglas cover.

learning. The continuity theorem holds the extreme view that this range comprises *all* discriminable cues impinging on the organism during reinforced trials. The discontinuity position holds the equally extreme view that this effective range comprises only *one* set of discriminable cues at a time, or two sets of cues in the event the organism is entertaining a couple of hypotheses simultaneously (12). The continuity position is a theorem derived from stimulus-response reinforcement theory. For reasons that are rather obscure to the present writers, the discontinuity position has often been stated as an integral part of cognitive theory. The question we wish to raise is whether it is necessary for *any* theory of learning to take an axiomatic position on what range of cues is being responded to in a given situation. It is our contention that this is an empirical question, and an important one.

Take stimulus-response reinforcement theory as a first case in point. Need one assume that reinforcement operates to strengthen response tendencies to *all* discriminable cues present in the situation? As Hull puts it, "According to the 'law of reinforcement' . . . , every one of the receptor discharges and receptor-discharge preservations active at the time that the to-be-conditioned reaction occurs must acquire an increment of habit strength" (9, p. 206). Yet it is known that rats in a jumping-stand situation learn more rapidly to respond to the cues close to the bottom of the card, near the edge to which they are jumping (e.g., Ehrenfreund, 8). And, indeed, it is also well known that a "set" to attend to certain kinds of objects or events in the environment leads to faster and better identification and learning of these than of so-called incidental material—although both kinds of materials are impinging on the organism's sensory apparatus (e.g., Chapman, 6; Postman and Senders, 20).

It will be argued, to be sure, that some stimulus-response theorists have taken the matter into account, the contention being supported by reference to the work of Ehrenfreund (8) and Spence (23). As Ehrenfreund remarks in commenting on the classical "one-cue-at-a-time" discontinuity position, "[continuity] theorists have emphasized the point that the conditions determining reception of the stimulus cues are extremely important" (8, pp. 417–418). He points out that since one must consider the conditions that determine what cues will be "picked up," many of the experiments of the continuity theorists have by-passed the problem by utilizing very simplified discrimination problems, preferring, for example, a simple black-white discrimination rather than pattern discrimination. In a pattern discrimination, "until the S learns the appropriate orienting act that will provide discriminably different retinal patterns of discrimination, the conditions necessary for the development of differential associations do not obtain" (p. 418). Presumably, such orienting acts involve "reception of specific cue aspects of the total stimulus complex on the area of the retina in which the ganglion cells are most dense" (p. 418). This sequence of orientation learning is what Spence calls "the acquisition of receptor exposure adjustments" (23, p. 720).

Indeed, Spence has recognized the importance of this problem. But to reduce the matter at this early stage of investigation to a question of adequate retinal stimulation seems to the present writers somewhat premature. For example, the 15° or 20° of visual arc that separates figures at the middle and at the bottom of a jumping-stand card, shown by Ehrenfreund (8) to matter so much in speed of acquisition of a pattern discrimination, can hardly make that much histological difference in ganglion density in the rat's eye—particularly in view of the amount of VTE behavior on a jumping stand. Moreover, it seems highly dubious that such a peripheralist explanation would account for what is known about "paying attention to cues" in human subjects. It seems to us that for the moment it is a hindrance to experimental ingenuity to rule out *central* factors that could operate to restrict the range of cues to which an animal is oriented in any given situation. Finally, there are a sufficient number of studies, such as those of Lashley (13) and Snygg (21, 22), to indicate that even when one cue is impinging on the *same* segment of the retina at which another cue is impinging, one may be learned and the other not. Lashly and Wade (14) have made much of this point in criticism of the "all-cues-are-relevant" position, and quite justifiably.

As for cognitive theory, a comparable argument applies. Why should it be axiomatic that while an organism is responding in terms of one cue, he is learning *nothing* about any other cues in the situation? The presence of a "hypothesis" or a systematic response tendency directed to one discriminandum does not necessarily blind the organism to all else. Indeed, the same kind of experiment can be used in evidence against this extreme view as is used against the "all-cues-are-relevant" viewpoint. The subjects in Chapman's experiment (6) *do* pick up features of the stimulus to which they have not been set. Learning of materials *does* occur without a set to learn, as in the Postman and Senders (20) experiment. And, indeed, there is an imposing array of evidence that while responding to one feature of a situation, animals learn about other features of the situation, as indicated by either transfer (e.g., Blum and Blum, 4) or by cue reversal (e.g., McCulloch and Pratt, 16; Ehrenfreund, 8; and various studies reviewed by Melton, 17 and Spence, 23). To hold the one-cue-at-a-time position is to deny, among other things, that there can be anything approximating a "general set" or "looking around" behavior—a highly restrictive premise for a cognitive theory of learning. Again, we would hold that the number of cues to which the organism attends is not a matter to be settled at the axiomatic level, but should be treated as an empirical problem. It is indeed the case that results by Lashley (13) and by Lashley and Wade (14) are seemingly contradictory to those of Blum and Blum (4), and that Krechevsky (12) can be pitted against McCulloch and Pratt (16). The point is not to establish one set of results *over* the other, but to discover what in one experiment leads organisms to operate on one cue at a time, and, in another, on several at a time.

Krechevsky (12) has, of course, remarked that animals can attend to two cues at a time, as judged from the fact that their behavior can be shown to be simultaneously systematic with respect to two cues. That is, animals could be responding above chance to both a visual and a spatial cue. But his position still implies that the animal is only learning about those aspects of the environment to which it is systematically responding. If it has a double hypothesis, then it is learning only about the two cues involved. We would hold that it is quite unnecessary to *assume* that an organism has to be responding systematically to a cue in order to learn about it—or, for that matter, to assume anything else. The matter is an empirical one.

In sum, we would take the position that a concept like *breadth of learning* would increase the richness of either a cognitive or an S-R theory of learning. By considering breadth of learning as a dependent variable, theorists of either stripe would be more readily disposed to ask the critical question, "What determines breadth of learning or the range of cues about which the organism picks up information in the course of a particular learning sequence?"

Tolman (27) has addressed himself to this problem in his paper on "Cognitive Maps in Rats and Men" by inquiring into the determinants of broad and narrow cognitive maps. For one of the determinants of a narrow or a broad map is the range of cues utilized by an organism during cognitive acquisition. It would appear that there are several conditions that have the effect of narrowing the range of cues. One of these is motivational level. Tolman suggests, in a manner which he describes as "brief, cavalier, and dogmatic," that excessive motivation or frustration narrows a cognitive map, i.e., reduces the breadth of learning. A better case than suggested by his three deprecatory adjectives can actually be made for his claim. For example, Johnson (10) has recently shown that the degree of learning of the position of an irrelevant reward depends upon the strength of the drive conditions operative during learning. Animals run in a T maze while moderately thirsty and satiated for food learned the position of the food better than animals who were satiated for food but extremely thirsty. Indeed, in the paper by Spence and Lippitt that initiated the Iowa studies of latent learning, the authors suggest that "latent learning does not occur in the situation where the animals 'perceived' the subsequent test goal object (food) while motivated for another goal object (water), but latent learning does occur where complete satiation made for no particular goal directedness in the experimental set-up during the training series" (24, p. 429). Subsequent studies indicate that such an "all-or-none" position is too strongly stated and that animals operating under drive condition can learn the location of an irrelevant reward (e.g., Deisenroth and Spence, 7; Strange, 25; Walker, Knotter, and DeValois, 29; as well as the Johnson study cited above). Thistlethwaite sums up the evidence by remarking that a reasonable inference from the Spence-Lippitt studies is that "the stronger the drive, the poorer the irrelevant-incentive learning" (26, p. 111). Melton (18), approaching the literature from a different conceptual position, reaches much the same conclusion. The Johnson experiments appear to confirm the generalization, they being the first study systematically to vary drive level in this type of experiment.

Another source of evidence comes from studies of incidental learning and "incidental attention" under conditions of increased incentive. Bahrick (1), for example, has shown that while intentional learning was faster when the incentive given his human subjects was increased, the amount of incidental learning decreased. In a similar vein, Bahrick, Fitts, and Rankin (2) have shown that increase in incentive leads to a higher degree of selective attention for those parts of a complex task that subjects interpret as more important, with a concomitant tendency to pay less attention to other features of the situation.

A second condition suggested by Tolman as a narrower of cognitive maps is "an overdose of repetition" (27, p. 261). When an organism has "overlearned" a particular discrimination or problem situation, the effect of this overlearning may be to restrict the range of cues to which the organism attends. Thus, in Luchins' experiment (15), subjects who were given much practice with one kind of solution of a problem are less likely to notice when the same kind of problem can be solved by simpler means. To be sure, the experimental situation employed by Luchins is of such complexity that the effects may be due to many other factors known to be operative in problem solving. In any case, it seems a likely hypothesis that overlearning may operate in the way described.

Another suggestive finding from which one may infer that mechanization has the effect of narrowing the range of cues responded to is reported by Walker, Knotter, and DeValois (29). Again, using the irrelevant-incentive experimental design with animals running under motivated rather than satiated conditions, they found that animals who developed "strong" position preferences learned less about the location of the irrelevant incentive than those with weak position preference. Thistlethwaite (26) notes that in at least three of the experiments reporting no learning of the position of an irrelevant incentive the animals had acquired strong positional preferences during initial training (Kendler, 11; Spence and Lippitt, 24; Walker, 28). We are interpreting strong positional preferences as indicative of mechanization of behavior of a sort also produced by overlearning.

AN ILLUSTRATIVE EXPERIMENT

Now let us return to the experiment mentioned briefly in the introductory paragraphs of this paper. Breadth of learning will be denoted operationally in this experiment as the learning of single alternation while responding systematically to door brightness. What might favor such "breadth of learning"? We propose to test two possible factors: strength of motivation and degree of overlearning. Highly motivated animals should show less breadth of learning than moderately motivated ones. Animals overtrained on the original black-white discrimination should be less likely to pick up the single-alternation cue when it is introduced. For good measure, highly motivated and overtrained animals should show the least breadth of learning in the sense of being least able to pick up and use the added single-alternation cue.

Method. The design of the experiment is readily summarized. Fifty-two mature male albino rats were divided into four experimental groups and four matching control groups, each group consisting of from six to eight animals. The experimental groups were all trained to certain levels of performance on a four-unit black-white discrimination apparatus to go either to black or to white. (See Fig. 1.) The correct doors, during this black-white learning, were always in a random pattern with respect to left-right position, with the restriction that the correct door was equally often on the right and the left on each trial. The animals were then given 20 trials of training in which the correctly colored doors were always in a single-alternation pattern, half of the animals going left-right-left-right, the other half right-left-right-left. The control groups received the same original black-white discrimination training on randomly positioned black and white doors and during the 20 critical trials were continued, as before, on the same random-position black-white discrimination. The experimental groups that received both black-white and single-alternation cues during the critical period we shall call the Double-Cue groups. The control animals we shall refer to as Single-Cue groups.

After the critical 20 trials, all animals were trained to criterion on a single-alternation pattern with uniform gray doors at each choice point in the four-unit apparatus.

The four experimental or Double-Cue groups, each with its corresponding control or Single-Cue group, were subjected to the following treatments:

1. *Learn–12-hr. deprived.* Thirty trials of black-white training before critical 20 trials. Twelve-hour food deprivation.
2. *Learn–36-hr. deprived.* Thirty trials of black-white training before critical 20 trials. Thirty-six–hour food deprivation.
3. *Overlearn–12-hr. deprived.* One hundred trials of black-white training before critical 20 trials. Twelve-hour food deprivation.
4. *Overlearn–36-hr. deprived.* One hundred trials of black-white training before critical 20 trials. Thirty-six–hour food deprivation.

The initial 30 trials of black-white training were sufficient to bring all animals to a criterion of 100 per cent on ten successive trials. This apparently small number of trials is not surprising, since the animals make four discriminations on each trial. Thus, the critical trials, whether single-cue or double-cue, were run in all instances with the animal making virtually no errors at all. Training on single alternation with gray doors was carried on at the same level of motivation used for the animal during black-white training and during the critical trials.

Animals were run ten trials per session and the sessions were separated by a 48-hr. period. After each trial, an animal was allowed to feed in the goal box on a rich wet mash for 10 to 15 sec. After running its ten trials, it was fed away from its home cage for 30 min. All animals were next fed 12 hr. later, again for 30 min. The harshly deprived group was not fed outside the apparatus again until the animals had completed their next set of ten trials 36 hr. later. Moderately deprived animals were fed once more before their next trials, 12 hr. later. Water was available to the rats at all times.

Results. Recall the first question. Will animals learn the single-alternation pattern during the critical period when they are responding effectively to door color? A comparison of pooled Double-Cue groups and pooled Single-Cue groups (Table 1) indicates that in the first ten trials of gray-door, single-alternation training Double-Cue groups made 8.0 errors and Single-Cue groups made 12.0 errors. An error, of course, is attempting to get through the wrong door in any unit. The difference in errors is significant at the .005 level. Thus we may conclude that the Double-Cue animals do pick up information about the "incidental cue." The finding conforms in general terms to those reported by Blum and Blum (4).

We come now to the test of the two factors that might limit breadth of learning. Consider degree of original learning of the black-white discrimination prior to the double-cue period. The Double-Cue animals that learned to criterion show a highly significant savings as compared to their Single-Cue controls (Table 1). The Double-Cue animals that underwent original overlearning do not do significantly better than their controls.

With respect to motivation, Double-Cue animals that operated under 12-hr. deprivation conditions do significantly better than their Single-Cue controls. But Double-Cue animals that operate under 36-hr. deprivation do not do better than their controls (Table 1).

Finally, treating each of the four conditions separately, we find the Double-Cue animals that learned to criterion and operated at 12 hr. of deprivation

TABLE 1

MEAN NUMBER OF ERRORS AND SAVINGS ON FIRST TEN TRIALS OF SINGLE ALTERNATION WITH GRAY DOORS

(Chance performance [50%] is 20 errors)

Group	Single Cue	Double Cue	Savings	t-Test Reliability*
All animals	12.0	8.0	4.0	<.005
Learn to criterion	12.9	7.4	5.5	<.005
Overlearn	10.9	8.7	2.3	N.S.
12-hr. deprived	12.6	7.8	4.8	.01
36-hr. deprived	11.4	8.2	3.2	N.S.
Learn–12 hr.	13.6	7.8	5.8	.025
Learn–36 hr.	12.3	6.8	5.5	<.05
Overlearn–12 hr.	11.5	7.8	3.7	N.S.
Overlearn–36 hr.	10.3	9.5	.8	N.S.

* Confidence levels are for a one-tailed test.

doing best in comparison with their controls; the overlearned, 36-hr. deprived animals do least well in comparison with their control group. An analysis of these group results suggests that overlearning has a more marked narrowing effect than high drive.

A word about the Single-Cue groups is in order. A comparison of the figures in Table 1 shows that overlearned animals learn a new task somewhat faster than ones with less prior training, a fact probably attributable to maze-wiseness. Similarly, highly motivated Single-Cue animals learn single alternation faster than moderately motivated ones. None of the differences among the Single-Cue groups is, however, significant.

Are there any observable features of the overt responses of animals who, during the critical period, pick up the alternation cue as compared with their less "observant" fellows? The first and most obvious difference to look for is in the number of vicarious-trial-and-error responses (VTE's) made by animals in those groups given the double cue during the critical trials. Table 2 contains such a comparison. Since the apparatus contains four choice points, the maximum number of VTE's possible during the 20 critical trials is 80, only one VTE being counted per choice point. Comparing the groups found before to be at the extremes on savings, we find that the group with the largest saving (Learn–12-hr. deprived) showed the greatest amount of VTE, while the group with the smallest saving (Overlearn–36-hr. deprived) showed the least VTE. The former made VTE responses on the average at half the choice points encountered, the latter at a quarter of the choice points. The other groups were between these in frequency of VTE. An analysis of variance of number of VTE's during the critical trials reveals that degree of motivation is significant at the .05 level, and amount of learning falls slightly short of significance at this confidence level.

Another possible difference between the groups is in speed of running during the 20 critical trials. Here the argument is somewhat ambiguous. In a visual task, it might be argued, speed might lead the animal to "overlook" cues. But the extra cue provided during the critical period was a positional one: single alternation. It might be argued that fast running would give better kinesthetic cues for picking up single alternation. In any case, the data on running speed are quite unrevealing. As might be expected, the more highly motivated animals ran the maze more quickly. On the other hand, those animals with initial overlearning

TABLE 2

AVERAGE NUMBER OF VTE RESPONSES DURING 20 CRITICAL TRIALS IN ANIMALS OF THE GROUPS GIVEN THE DOUBLE-CUE TREATMENT

(Maximum number of VTE's possible is 80)

Group	No. of VTE's
Learn–12-hr. deprived	40.2
Learn–36-hr. deprived	27.8
Overlearn–12-hr. deprived	29.5
Overlearn–36-hr. deprived	21.8
Combined Learn	34.0
Combined Overlearn	25.6
Combined 12-hr. deprived	34.9
Combined 36-hr. deprived	24.8

of the black-white discrimination were slower than the group with initial learning to criterion. Thus, one factor associated with failure to pick up the extra cue (high motivation) is associated with fast running, the other (overlearning) with slow running.

DISCUSSION AND CONCLUSION

We would propose as a first and most general conclusion that there are conditions that lower the likelihood that an organism will attend to or acquire knowledge about environmental cues not immediately relevant to the to-be-reinforced responses he is making. Breadth of learning can fruitfully be treated as a dependent variable, and two conditions affecting it have been proposed: overlearning and overmotivation.

Beyond this, the question arises as to the nature of the effect: What mediates breadth of learning? In an earlier section, the proposal was made that a peripheralist explanation in terms of the organism orienting his relevant receptors to the stimulus would be premature. In the present experiment, the rats were peripherally exposed to the second or extra cue that they were to be tested on later. At least there is no basis for saying that the alternating kinesthetic pattern was not registering on appropriate receptors. A glance at Fig. 1 suffices to recognize that the apparatus was designed in a way to guarantee weaving about. Our conclusion must be that a peripheral sensory stimulus may or may not be utilized in discrimination learning, depending upon the operation of certain internal processes. In view of this, the axiomatic position of continuity theory that *all* cues present when a response is reinforced are relevant seems no more justified than the axiomatic position of discontinuity theory that only the *one* cue toward which the organism has a hypothesis is relevant.

What of the effect of motivation and mechanization on breadth of learning? Information analyses of recognition and identification phenomena have made it clear that the larger the number of alternative cues for which the organism is set, the greater the difficulty involved in recognizing any one of them upon appearance. Thus, Miller, Heise, and Lichten (19) and Bruner, Miller, and Zimmerman (5) have shown that the larger the number of alternative words expected in an auditory articulation test, the greater the signal-to-noise ratio necessary in order for each of the same words to be recognized. Miller, Postman, and Bruner (18) have shown that an increase in independent alternatives requires an increase in exposure duration for visual identification to occur.

Now, the more pressing the requirement that an organism reach a goal rapidly, the more hindering will be a set for considering many alternative cues. Thus, to speed up goal attainment, an organism sacrifices breadth of attention and consequently breadth of learning.

Extreme motivation impelling an organism to the attainment of a specific goal may provide almost a paradigm of the kind of adjustive requirement that compels speed and efficiency with respect to minimal cues. Thus, we can summarize the matter by saying that strong motivation has the effect of speeding up learning at the cost of narrowing it.

As for "a dose of overlearning," as Tolman (27) calls one of the conditions for producing a narrow cognitive map, its effect may also be to restrict cue utilization in a given situation so that new cues introduced after stabilization has occurred are not as likely to be utilized. Why or how the range of cues utilized should be restricted in overlearning is not clear. There is fair evidence that it is. Generalization gradients under repeated classical conditioning narrow (3); and it is a matter of common sense observation that when performance of a particular task becomes mechanized, we "fall into a set" and may fail to notice that the situation has changed, as in the Luchins experiment mentioned earlier (15).[2]

By way of general conclusion, three points deserve emphasis:

1. It is fruitless to take an axiomatic position about the range of cues that an organism will utilize in making a discrimination. The range of cues utilized, or attended to, or associated with a response, is a function of determinate processes and should be treated empirically as a dependent variable. We have called this variable *breadth of learning*.

2. Conditions requiring increased speed and efficiency of goal attainment by an organism may have the effect of narrowing the range of environmental cues to which the organism responds. High motivation and intensive practice are known to increase the specific efficiency of learning. Yet these are the very conditions that appear to narrow the breadth of learning. In short, the picture given by such traditional measures as latency, speed, and even errors fails to encompass a critical aspect of learning: how much about the environment is the organism picking up over and beyond what is required by the task at hand?

3. If the variable we have called breadth of learning is to take its place beside other traditional learning variables such as those mentioned above, it is necessary to devise experimental paradigms for its measurement. Two such paradigms already exist. The first is the "cue reduction" procedure in which an array of relevant cues is provided in the original learning and, in later test situations, is systematically reduced in number. A more efficient variant is to establish discrimination on the basis of one set of cues and then to introduce additional cues, finally removing the original cues in the testing situation. This procedure has been used by Lashley (13), Blum and Blum (4), Greenbaum,[3] and by us in the present experiment. A third technique involves devising situations in which the actual utilization of cues can be observed. Counting the number of VTE responses is one way of approximating this procedure. Thus, in the present experiment, it was found that the moderately motivated animals and those not receiving overlearning trials showed the greatest amount of VTE during the critical trials when both cues were present. Wyckoff (30) has recently introduced an ingenious and direct way of observing cue utilization. Using the homogeneous environment of a Skinner box, he required of his pigeons that they press on a lever in order to have the relevant discriminanda appear. Without these discriminanda the birds could not press the food-delivery lever at a level better than chance. Wyckoff was able, by this means, to study the growth of single-cue utilization under a variety of conditions. It would be relatively easy to extend the technique to a multiple-cue experiment.

Finally, and almost as a postscript, it seems somehow ironic that psychologists for the most part have neglected an aspect of learning—call it breadth, perspective, or what you will—that is so much a feature of our instructional lives. We constantly seek to create the conditions in which our students will grasp the many and varied aspects of the phenomena to which we are introducing them. Yet in our theories of learning we have taken the side of the hedgehog in the classic line, "The fox knows many things, but the hedgehog knows one thing."

---

[2] On the other hand, it may be pointed out that mechanization of a performance may have the effect of "freeing" the individual to notice those aspects of his environment not directly related to his task—may indeed broaden learning. The highly skilled skier, for example, is more likely to notice features of the landscape not related to his immediate task of descending a slope than is the novice—assuming equal motivation. There is obviously an additional variable, or variables, operating in overlearning that is still largely unknown to us.

[3] Greenbaum, J. Personal communication, 1953.

REFERENCES

1. BAHRICK, H. P. Incidental learning under two incentive conditions. *J. exp. Psychol.*, 1954, **47**, 170–172.
2. BAHRICK, H. P., FITTS, P. M., & RANKIN, R. R. Effect of incentives upon reactions to peripheral stimuli. *J. exp. Psychol.*, 1952, **44**, 400–406.
3. BERITOFF, J. S. On the fundamental nervous processes in the cortex of the cerebral hemispheres. I. The principle stages of the development of the individual reflex: its generalization and differentiation. *Brain*, 1924, **47**, 109–148.
4. BLUM, R. A., & BLUM, JOSEPHINE S. Factual issues in the continuity controversy. *Psychol. Rev.*, 1949, **56**, 33–50.
5. BRUNER, J. S., MILLER, G. A., & ZIMMERMAN, C. Discriminative skill and discriminative matching in perceptual identification. *J. exp. Psychol.*, in press.
6. CHAPMAN, D. W. Relative effects of determinate and indeterminate *Aufgaben*. *Amer. J. Psychol.*, 1932, **44**, 163–174.
7. DEISENROTH, C. F., & SPENCE, K. W. An investigation of learning in the white rat. *Psychol. Bull.*, 1941, **38**, 706. (Abstract)
8. EHRENFREUND, D. An experimental test of the continuity theory of discrimination with pattern vision. *J. comp. physiol. Psychol.*, 1948, **41**, 408–422.
9. HULL, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
10. JOHNSON, E. E. The role of motivational strength in latent learning. *J. comp. physiol. Psychol.*, 1953, **45**, 526–530.
11. KENDLER, H. H. An investigation of latent learning in a T-maze. *J. comp. physiol. Psychol.*, 1947, **40**, 265–270.
12. KRECHEVSKY, I. A study of the continuity of the problem solving process. *Psychol. Rev.*, 1938, **45**, 107–133.
13. LASHLEY, K. S. An examination of the "continuity" theory as applied to discrimination. *J. gen. Psychol.*, 1942, **26**, 241–265.
14. LASHLEY, K. S., & WADE, M. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.
15. LUCHINS, A. S. Mechanization in problem solving: the effect of *Einstellung*. *Psychol. Monogr.*, 1942, **54**, No. 6 (Whole No. 248).
16. MCCULLOCH, T. L., & PRATT, J. G. A study of the presolution period in weight discrimination by white rats. *J. comp. Psychol.*, 1934, **18**, 271–290.
17. MELTON, A. W. Learning. *Annu. Rev. Psychol.*, 1950, **1**, 9–30.
18. MILLER, G. A., BRUNER, J. S., & POSTMAN, L. Familiarity of letter sequences and tachistoscopic identification. *J. gen. Psychol.*, in press.
19. MILLER, G. A., HEISE, G. A., & LICHTEN, W. The intelligibility of speech as a function of the context of the test materials. *J. exp. Psychol.*, 1951, **41**, 329–335.
20. POSTMAN, L., & SENDERS, VIRGINIA L. Incidental learning and generality of set. *J. exp. Psychol.*, 1946, **36**, 153–165.
21. SNYGG, D. The relative difficulty of mechanically equivalent tasks: I. Human learning. *J. genet. Psychol.*, 1935, **47**, 299–320.
22. SNYGG, D. The relative difficulty of mechanically equivalent tasks: II. Animal learning. *J. genet. Psychol.*, 1935, **47**, 321–336.
23. SPENCE, K. W. Theoretical interpretations of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 690–729.
24. SPENCE, K. W., & LIPPITT, R. "Latent" learning of a simple maze problem with relevant needs satiated. *Psychol. Bull.*, 1940, **37**, 429. (Abstract) (Also reported in Spence, K. W., & Lippitt, R. An experimental test of the sign-gestalt theory of trial and error learning. *J. exp. Psychol.*, 1946, **36**, 491–502.)
25. STRANGE, J. R. Latent learning under conditions of high motivation. *J. comp. physiol. Psychol.*, 1950, **43**, 194–197.
26. THISTLETHWAITE, D. A critical review of latent learning and related experiments. *Psychol. Bull.*, 1951, **48**, 97–129.
27. TOLMAN, E. C. Cognitive maps in rats and men. *Psychol. Rev.*, 1948, **55**, 189–208.
28. WALKER, E. L. Drive specificity and learning. *J. exp. Psychol.*, 1948, **38**, 39–49.
29. WALKER, E. L., KNOTTER, M. C., & DEVALOIS, R. L. Drive specificity and learning: the acquisition of a spatial response to food under conditions of water deprivation and food satiation. *J. exp. Psychol.*, 1950, **40**, 161–168.
30. WYCKOFF, L. B. The role of observing responses in discrimination learning. *Psychol. Rev.*, 1952, **59**, 431–442.

(Received January 18, 1954)

Psychological Review
Vol. 62, No. 1, 1955

A THEORY OF DISCRIMINATION LEARNING[1]

FRANK RESTLE

Stanford University[2]

This paper presents a theory of two-choice discrimination learning. Though similar in form to earlier theories of simple learning by Estes (5) and Bush and Mosteller (2,3), this system introduces a powerful new assumption which makes definite quantitative predictions easier to obtain and test. Several such predictions dealing with learning and transfer are derived from the theory and tested against empirical data.

The stimulus situation facing a subject in a trial of discrimination learning is thought of as a set of cues. A subset of these cues may correspond to any thing—concrete or abstract, present, past, or future, of any description—to which the subject can learn to make a differential response. In this definition it does not matter whether the subject actually makes a differential response to the set of cues as long as he has the capacity to learn one. An individual cue is thought of as "indivisible" in the sense that different responses cannot be learned to different parts of it. Informally, the term "cue" will occasionally be used to refer to any set of cues, all of which are manipulated in the same way during a whole experiment.

In problems to be analyzed by this theory, every individual cue is either "relevant" or "irrelevant." A cue is relevant if it can be used by the subject to predict where or how reward is to be obtained. For example, if food is always found behind a black card in a rat experiment, then cues aroused by the black card are relevant. A cue aroused by an object uncorrelated with reward is "irrelevant." For example, if the reward is always behind the black card but the black card is randomly moved from left to right, then "position" cues are irrelevant. These concepts are discussed by Lawrence (6).

In experiments to be considered, the subject has just two choice responses. No other activities are considered in testing the theory. Any consistent method of describing these two responses which can be applied throughout a complete experiment is acceptable in using this theory.

THEORY

In solving a two-choice discrimination problem the subject learns to relate his responses correctly to the relevant cues. At the same time his responses become independent of the irrelevant cues. These two aspects of discrimination learning are represented by two hypothesized processes, "conditioning" and "adaptation."

Intuitively, a conditioned cue is one which the subject knows how to use in getting reward. If k is a relevant cue and $c(k,n)$ is the probability that k has been conditioned at the beginning of the nth trial, then

$$c(k,n+1)=c(k,n)+\theta[1-c(k,n)] \quad [1]$$

---

[1] This paper is adapted from part of a Ph.D. dissertation submitted to Stanford University. The author is especially indebted to Dr. Douglas H. Lawrence and to Dr. Patrick Suppes for encouragement and criticism. Thanks are also due Dr. W. K. Estes who loaned prepublication manuscripts and Dr. R. R. Bush who pointed out some relations between the present theory and the Bush-Mosteller model (3).

[2] Now at the Human Resources Research Office, The George Washington University.

is the probability that it will be conditioned by the beginning of the next trial. On each trial of a given problem a constant proportion, θ, of unconditioned relevant cues becomes conditioned.

To the extent that a conditioned cue affects performance, it contributes to a correct response only, whereas an unconditioned relevant cue contributes equally to a correct and to an incorrect response.

Intuitively, an adapted cue is one which the subject does not consider in deciding upon his choice response. If a cue is thought of as a "possible solution" to the problem, an adapted cue is a possible solution which the subject rejects or ignores. If $a(k,n)$ is the probability that irrelevant cue k has been adapted at the beginning of the nth trial, then

$$a(k,n+1) = a(k,n) + \theta[1 - a(k,n)] \quad [2]$$

is the probability that it will be adapted by the beginning of the next trial. On each trial of a given problem a constant proportion of unadapted irrelevant cues becomes adapted. An adapted cue is nonfunctional in the sense that it contributes neither to a correct nor to an incorrect response.

It will be noticed that the same constant θ appears in both equations 1 and 2. The *fundamental simplifying assumption* of this theory deals with θ. This assumption is that

$$\theta = \frac{r}{r+i}, \quad [3]$$

where r is the number of relevant cues in the problem and i is the number of irrelevant cues. Thus, θ is the proportion of relevant cues in the problem. This proportion is the same as the fraction of unconditioned cues conditioned on each trial, and the fraction of unadapted cues adapted on each trial.

The performance function $p(n)$, representing the probability of a correct response on the nth trial, is in accord with the definitions of conditioning and adapting given above. The function is in the form of a ratio, with the total number of unadapted cues in the denominator and the number of conditioned cues plus one-half times the number of other cues in the numerator. Thus conditioned cues contribute their whole effect toward a correct response, adapted cues contribute nothing toward either response, and other cues contribute their effect equally toward correct and incorrect responses. Formally,

$$p(n) = \frac{\sum\limits^{r} c(k,n) + \frac{1}{2}\sum\limits^{r} [1 - c(k,n)] + \frac{1}{2}\sum\limits^{i} [1 - a(k,n)]}{r + \sum\limits^{i} [1 - a(k,n)]}. \quad [4]$$

Here $\sum\limits^{r}$ is the sum taken over the r relevant cues and $\sum\limits^{i}$ is the sum taken over the i irrelevant cues.

SOME CONSEQUENCES REGARDING SIMPLE LEARNING

If the subject is naive at the beginning of training, so that for any relevant cue k, $c(k,1) = 0$, and for any irrelevant cue k, $a(k,1) = 0$, and if he receives n trials on a given problem, then by mathematical induction it can be shown that if k is relevant,

$$c(k,n+1) = 1 - (1-\theta)^n \quad [5]$$

and if k is irrelevant,

$$a(k,n+1) = 1 - (1-\theta)^n. \quad [6]$$

Under these circumstances we can substitute equations 5 and 6 into equation 4 and, taking advantage of the simplifying effects of equation 3, we have

$$p(n) = 1 - \tfrac{1}{2}\frac{(1-\theta)^{n-1}}{\theta + (1-\theta)^n}. \quad [7]$$

Plotting equation 7 shows that p is an **S**-shaped function of n with an asymptote (for $\theta > 0$) at 1.00. Also, $p(1) = \tfrac{1}{2}$. Since $p(n)$ is a monotonic increasing function of θ we can estimate θ from observations of performance. If we want to know the theoretical proportion of relevant cues in a problem for a particular subject, we have the subject work on the problem, record his performance curve, and solve equation 7 for θ. This result depends directly upon the simplifying assumption of equation 3.

Since the instability of individual learning curves makes it difficult to fit curves to them, it is fortunate that θ can be determined in a different way. Suppose a subject makes E errors in the course of solving the problem to a very rigorous criterion and it is assumed for practical purposes that he has made all the errors he is going to make. Theoretically, the total number of errors made on a problem can be written

$$E = \sum_{n=1}^{\infty} [1 - p(n)].$$

Under the conditions satisfying equation 7, this can be evaluated approximately by using the continuous time variable t in place of the discrete trial variable n, and integrating. The result of this integration is that

$$E \simeq \tfrac{1}{2} + \tfrac{1}{2}\frac{\log\theta}{(1-\theta)\log(1-\theta)}. \quad [8]$$

By equation 8, which relates the total number of errors made on a problem to θ, it is possible to make relatively stable estimates of θ.

AN EMPIRICAL TEST OF THE SIMPLE LEARNING THEORY—COMBINATION OF CUES

Consider three problems, s_1, s_2, and s_3, all of which involve the same irrelevant cues. Two of the problems, s_1 and s_2, have entirely separate and different relevant cues, while in problem s_3 all the relevant cues of s_1 and s_2 are present and relevant. That is, $r_3 = r_1 + r_2$ and $i_1 = i_2 = i_3$. If we know θ_1 and θ_2 we can compute θ_3, since by equation 3

$$\theta_1 = r_1/(r_1 + i)$$
$$\theta_2 = r_2/(r_2 + i)$$
$$\theta_3 = (r_1 + r_2)/(r_1 + r_2 + i).$$

Solving these equations for θ_3 in terms of θ_1 and θ_2 we get

$$\theta_3 = (1-\theta_1)(1-\theta_2)/(1-\theta_1\theta_2). \quad [9]$$

This theorem answers the following question: Suppose we know how many errors are made in learning to use differential cue X and how many are used to learn cue Y, then how many errors will be made in learning a problem in which either X or Y can be used (if X and Y are entirely discrete)?

Eninger (4) has run an experiment which tests equation 9. Three groups of white rats were run in a **T** maze on successive discrimination problems. The first group learned a visual discrimination, *black-white*, the second group learned an auditory discrimination, *tone–no-tone*, and the third group had both cues available and relevant.

Since each group was run to a rigorous criterion, total error scores are used to estimate θ_1 and θ_2 by equation 8.[3] The values estimated are

---

[3] Total error scores do not appear in Eninger's original publication and are no longer known. However, trials-to-criterion scores were reported. Total error scores were

$\theta_1 = .020$, based on an estimated average of 98.5 errors made on the auditory-cue problem, and $\theta_2 = .029$, based on an estimated average of 64.5 errors on the visual-cue problem. Putting these two values into equation 9 we get

$$\theta_3 = .029 + .020 + 2(.020)(.029)/1 - (.020)(.029) = .049.$$

This value of θ_3 substituted into equation 8 leads to the expectation of about 33 total errors on the combined cues problem. In fact, an average of 26 errors was made by the four subjects on this problem. The prediction is not very accurate. However, only 14 animals were employed in the entire experiment, in groups of five, five, and four. Individual differences among animals within groups were considerable. If account is taken of sampling variability of the two single-cue groups and of the combined-cue group of subjects, the prediction is not significantly wrong. Further experimentation is needed to determine whether the proposed law is tenable.

It is easily seen that θ_3 will always be larger than θ_1 or θ_2 if all three problems are solved. Learning will always be faster in the combined-cues problem. Eninger (4) in his paper points out that this qualitative statement is a consequence of Spence's theory of discrimination. However, Spence's theory gives no quantitative law.

TRANSFER OF TRAINING

In order to apply this theory to transfer-of-training experiments in which more than one problem is used, certain assumptions are made. It is assumed that if a cue is conditioned in one problem and appears immediately thereafter as a relevant cue in a new problem, it is still conditioned. Likewise, an adapted cue appearing as an irrelevant cue in a new problem is adapted. However, if a conditioned cue is made irrelevant it is obviously no longer conditioned, since it cannot serve as a predictor of reward. Similarly, it is assumed that if an adapted cue is made relevant in a new problem, it becomes unadapted and available for conditioning.

According to the present definition of conditioning, a conditioned cue contributes to a correct response. Therefore the above assumptions will not hold if the relation between a cue and a reward is reversed in changing the problem. This theory cannot be used to analyze reversal learning, and is applicable only in cases in which relevant cues maintain an unchanging significance.

If two problems are run under the same conditions and in the same apparatus, and differ only in the degree of difference between the discriminanda (as where one problem is a *black-white* and the other a *dark gray-light gray* discrimination), it is assumed that both problems involve the same cues; but the greater the difference to be discriminated, the more cues are relevant and the less are irrelevant.

EMPIRICAL TESTS OF THE TRANSFER-OF-TRAINING THEORY

As Lawrence (7) has pointed out, it seems that a difficult discrimination is more easily established if the subjects are first trained on an easy problem of the same type than if all training is given directly on the difficult discrimination. The experimental evidence on this point raises the question of predicting transfer per-

estimated from trials-to-criterion scores by using other, comparable data collected by Amsel (1). Dr. Amsel provided detailed results in a personal communication.

formance from one problem to another, where the two problems involve the same stimulus dimension but differ in difficulty.

Suppose that problems s_1 and s_2 both require a discrimination along the same stimulus dimension and differ only in that s_2 is more difficult than s_1. Let θ_1 be the proportion of relevant cues in problem s_1 and θ_2 be the proportion of relevant cues in s_2. Suppose that the training schedule involves n trials on problem s_1 followed by j trials on problem s_2. Then the probability of a correct response on trial $n + j$ is

$$p(n+j) = \frac{\theta_2 + \frac{1}{2}(1-\theta_2)^{j-1}[\theta_1 - \theta_2 + (1-\theta_1)^n(1-\theta_1-\theta_2)]}{\theta_2 + (1-\theta_2)^{j-1}[\theta_1 - \theta_2 + (1-\theta_1)^{n+1}]}. \qquad [10]$$ [4]

This theorem can be tested against the results of experiments reported by Lawrence (7). He trained white rats in one brightness discrimination and transferred them to a more difficult problem for further training. A control group, which Lawrence called "HDG," learned the hard test problem without work on any other problem. The performance of this control group is used to estimate θ_2, the proportion of relevant cues in the test problem. The value found was .04.[5] Since the experimental subjects first worked on the pretraining problem without prior experience, their performance on the first problem serves to estimate θ_1, the proportion of relevant cues in the easier pretraining problem. Lawrence replicated the experiment, having two experimental groups, ATG No. 1 and ATG No. 2, each of which transferred abruptly from an easy pretraining problem to the test problem. Group ATG No. 1 had a very easy problem for which we estimate $\theta_1 = .14$. Group ATG No. 2 had a more difficult problem for which $\theta_1' = .07$.

For group ATG No. 1, $\theta_1 = .14$, $\theta_2 = .04$, and $n = 30$ since thirty trials of pretraining were given. From this information we can compute $p(n + j)$ for all j, using equation 10. The predicted transfer performance is compared with observed performance in Table 1. For group ATG No. 2, $\theta_1' = .07$, $\theta_2 = .04$, and $n = 50$ since fifty trials of pretraining were given. Here also, $p(n + j)$ can be computed. Prediction is compared with observed performance in Table 1, from which it can be seen that the predictions are

TABLE 1

Prediction of Easy-to-Hard Transfer in Rats*

Trials of Transfer Training	Proportion of Correct Responses			
	Group ATG 1		Group ATG 2	
	Observed	Predicted	Observed	Predicted
1–10	.66	.63	.81	.71
11–20	.70	.68	.83	.77
21–30	.74	.72	.81	.81
31–40	.84	.78		
41–50	.86	.83		

* Data from Lawrence (7).

---

[4] The justification of equation 10 involves no mathematical difficulties. On the first trial of transfer we know the probability that any cue relevant in the second problem is conditioned, since all cues relevant in the second problem were relevant in the first. Similarly, we know the probability that i_1 of the i_2 irrelevant cues are adapted. The other $i_2 - i_1$ cues are unadapted. Equations 1 and 2 can be applied at this point, and all terms divided by $r_1 + i_1 (= r_2 + i_2)$.

[5] These estimates were made by the unsatisfactory method of fitting equation 7 to group average learning curves. Therefore the results regarding Lawrence's experiment are approximate.

relatively accurate, though performance is higher than predicted.

Lawrence also considered the possibility that a gradual transition from easy through successively harder problems would result in rapid mastery of the difficult problem. He tested this proposition by giving another group of subjects a series of three pretest problems before the final test problem. The problems in order of ease of learning were, first, the problem learned by ATG No. 1 with $\theta_1 = .14$, an intermediate problem which was not otherwise used, the difficult pretest problem with $\theta_3 = .07$, and finally the test problem with $\theta_4 = .04$.

To estimate θ_2 in Lawrence's experiment where problem s_2 never was used separately in simple learning, we notice the relation of θ to differences between discriminanda in apparent foot-candles for problems s_1, s_3, and s_4 whose θ values are known. We know that if the problems are properly controlled, and the stimulus difference is zero foot-candles, there are no relevant cues and θ is zero. It was found that this assumption, along with available data, made it possible to write a tentative empirical function relating θ to the difference between discriminanda in foot-candles. This equation presumably holds only in the case of Lawrence's apparatus, training procedure, subjects, etc. The equation adopted is

$$\theta = .0988 \log_{10}(.4\, d) \qquad [11]$$

where d is the difference between discriminanda in foot-candles. It is emphasized that this equation has no theoretical significance and is merely expedient. From equation 11 it is possible to determine the θ value of the intermediate pretraining problem by interpolation. Table 2 gives the data and results of this interpolation.

TABLE 2

THE RELATION OF "DIFFERENCE BETWEEN STIMULI" AND θ VALUE OF PROBLEM*

Difference Between Discriminanda in Apparent Foot-Candles	Corresponding θ Value of Problem
67.7	.14
35.2	.113**
14.0	.07
5.9	.04
0.0	.00†

* Data from Lawrence (7).
** Estimated by interpolation from empirical equation 16.
† Theoretical—see text for explanation.

TABLE 3

PREDICTION OF TRANSFER PERFORMANCE OF RATS AFTER A SERIES OF PRETRAINING PROBLEMS*

Trials Working on Final Test Problem	Proportion of Correct Responses	
	Observed	Predicted
1–10	.73	.73
11–20	.82	.79
21–30	.87	.84
31–40	.89	.87
41–50	.90	.90

* Data from Lawrence (7).

Ten trials were given on each of the first three problems and fifty trials on the final test problem. Using the θ values in Table 2 it is possible to predict the test problem performance of subjects who have gone through gradual transition pretraining.[6] This prediction is compared with observed performance in Table 3. It may be noted that the correspondence between prediction and observation is in this case very close. Again, however, the prediction is consistently a little lower than observed performance.

---

[6] The general prediction for transfer through a series of problems which get successively more difficult can be derived by following through and repeating the reasoning in footnote 4. Since the resulting equations are extremely large and can be derived rather easily, they are not given here.

NEW DATA

The theory has thus far been tested against the behavior of rats. Its generality is now tested with college students in a simple discrimination learning task.

Subjects and procedure. The subjects in this experiment were 23 students in the elementary psychology course at Stanford University. The *S* was seated at one end of a table and told that his responses could be either "*A*" or "*B*". On each trial *S* saw a single stimulus, which was a black square on a circular white background. The two squares used on alternate trials differed in size. In problem s_1 the squares differed in height by $\frac{1}{4}$ in., in problem s_2 they differed by $\frac{1}{8}$ in. The mean height of each pair of squares was 3 in. The squares were viewed at a distance of about 6 ft.

For half the *S*s in each experimental group, the problem was to say "*A*" to the smaller square and "*B*" to the larger one. The other *S*s had the converse problem. The *S* was never told that the problem was a size discrimination. Stimuli were alternated randomly. A rest period was called after each ten trials and *S* was asked what he thought the correct solution to the problem was, and to outline possible solutions which had occurred to him. This method of questioning is a modification of Prentice's method (8).

Twelve *S*s were trained first on problem s_1 to a criterion of 15 successive correct responses and then transferred to problem s_2 and run to the same criterion. These *S*s made up the "Easy-Hard Transfer Group" called *EH*. The other 11 *S*s were trained first on s_2 and then transferred to s_1. This was the "Hard-Easy Transfer Group" called *HE*. The two groups were approximately equated for age, sex, and known special visual skills.

Results. Using the pretraining performance of the *EH* group, the average proportion of relevant cues, θ_1, was estimated at .254 by equation 8. Using the pretraining performance of the *HE* group, the average proportion of relevant cues in problem s_2 was estimated at $\theta_2 = .138$.

The transfer performance of group *EH*, which first learned the easy and then the hard problem, is predictable by equation 10. Since these subjects worked to a high criterion in pretraining, we can assume that $p(n)$ is negligibly different from one at the end of pretraining. Then by equation 7 we see that $(1-\theta_1)^{n-1}$ is small, and equation 10 simplifies to

$$p(n+j)=\frac{\theta_2+\frac{1}{2}(1-\theta_2)^{j-1}(\theta_1-\theta_2)}{\theta_2+(1-\theta_2)^{j-1}(\theta_1-\theta_2)}. \quad [12]$$

This theoretical function of j is compared with observed transfer performance in Table 4. It is seen that the correspondence is quite close with a negligible constant error.

This prediction is based on the formula which also predicted Lawrence's rat data. This confirmation suggests that the law can be applied to human as well as rat performance on this type of task.

Using the line of reasoning which developed equation 10 we can produce an equation to predict transfer performance from hard to easier problems of the same sort. Certain cues are relevant in the easy problem which were irrelevant in the harder one. These cues cannot be identified in the hard problem. For performance to be perfect in the easier problem all relevant cues must be identified. Therefore, when the subject transfers from the hard to the easier

TABLE 4

PREDICTION OF TRANSFER OF TRAINING FROM EASIER TO HARDER PROBLEM IN HUMAN SUBJECTS

Trials after Transfer to Second Problem	Proportion of Correct Responses	
	Observed	Predicted
1–5	.817	.821
6–10	.933	.895
11–15	.926	.941
16–20	.933	.966
21–25	.966	.988
26–30	.983	.994
31–35	1.000	1.000

TABLE 5

PREDICTION OF TRANSFER OF TRAINING FROM HARDER TO EASIER PROBLEM IN HUMAN SUBJECTS

Trials After Transfer to Second Problem	Proportion of Correct Responses	
	Observed	Predicted
1–4	.932	.883
5–8	.955	.960
9–12	.955	.984
13–16	1.000	.995

problem we should expect some small number of errors to be made. On the assumption that the hard problem was completely learned in pretraining, the formula for transfer performance on the easy problem is

$$p(n+j)=\frac{\theta_2+(\theta_1-\theta_2)(1-\theta_1)^{j-1}}{\theta_1} \quad [13]$$

where θ_1 is the proportion of relevant cues in the easy problem and θ_2 is the proportion of relevant cues in the harder problem. The proof of this theorem is similar to that of equation 12 above, and is not given here.

Equation 13 yields the prediction for transfer performance of the *HE* subjects. In Table 5 the prediction is compared with observed transfer performance.

Despite the very small frequencies predicted and observed, the prediction is quite accurate. In all, seven errors were made by eleven subjects, whereas a total of eight were expected. This is an average of .64 errors per subject observed, and .73 predicted.

DISCUSSION

The definition of a "cue" in terms of possible responses is selected because the theoretical results do not depend critically upon the nature of the stimulating agent. While cues are thought of as stimulus elements, these elements need not be of the nature of "points of color" or "elementary tones." If a subject can learn a consistent response to a certain configuration despite changes in its constituents, then the configuration is by definition a cue separate from its constituents. The intention is to accept any cue which can be demonstrated to be a possible basis for a differential response.

The process of conditioning described in this paper is formally similar to the processes of conditioning of Estes (**5**) and Bush and Mosteller (**2**,**3**). In the present theory conditioning takes place at each trial, not only on "reinforced" trials. In earlier theories conditioning is said to occur only on such reinforced trials. In two-choice discrimination the incorrect response has a high initial probability (one-half) because of the nature of the physical situation and the way of recording responses. Therefore, a theory of two-choice learning must account for the consistent weakening of such responses through consistent nonreinforcement.

The notion of adaptation used here is formally analogous to the operation of Bush and Mosteller's Discrimination Operator "*D*" (**3**). However, whereas Bush and Mosteller's operator is applied only on trials in which the reward condition is reversed for a cue, the present theory indicates that this process takes place each trial. In addition, while the Discrimination Operator and the process of adaptation are both exponential in form, Bush and Mosteller introduce a new exponential constant *k* for this purpose and the present theory uses the conditioning constant θ.

The major point differentiating the present theory from similar earlier theories is the use of the strong simplifying assumption identifying the exponential constant θ with the proportion of relevant cues. This assumption may appear intuitively unlikely, but if it should be shown by further experiment to be tenable, the predictive power of discrimination learning theory is enhanced. There seems to be no reason for abandoning so useful an assumption unless experimental results require it.

SUMMARY

A theory of two-choice discrimination learning has been presented. The theory is formally similar to earlier theories of Estes (5) and Bush and Mosteller (3) but differs somewhat in basic concepts and uses a new simplifying assumption.

From this theory three empirical laws are derived: one dealing with the combination of relevant cues, and two dealing with a special type of transfer of training. These laws permitted quantitative predictions of the behavior of four groups of rats and two groups of human subjects. Five of these six predictions were quite accurate, and the sixth was within the range of reasonable sampling deviation.

REFERENCES

1. AMSEL, A. Rate of learning a visual brightness discrimination as a function of discriminanda durations. *J. comp. physiol. Psychol.*, 1952, **45**, 341–346.
2. BUSH, R. R., & MOSTELLER, F. A mathematical model for simple learning. *Psychol. Rev.*, 1951, **58**, 313–323.
3. BUSH, R. R., & MOSTELLER, F. A model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58**, 413–423.
4. ENINGER, M. U. Habit summation in a selective learning problem. *J. comp. physiol. Psychol.*, 1952, **45**, 511–516.
5. ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
6. LAWRENCE, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40**, 175–188.
7. LAWRENCE, D. H. The transfer of a discrimination along a continuum. *J. comp. physiol. Psychol.*, 1952, **45**, 511–516.
8. PRENTICE, W. C. H. Continuity in human learning. *J. exp. Psychol.*, 1949, **39**, 187–194.

(Received January 14, 1954)

Psychological Review
Vol. 62, No. 1, 1955

A STIMULUS–RESPONSE ANALYSIS OF THE INTERACTION OF CUE–PRODUCING AND INSTRUMENTAL RESPONSES[1]

ALBERT E. GOSS

University of Massachusetts

Hull (8), Gibson (6), Dollard and Miller (3), and others (e.g., 1, 2, 5, 13) have been concerned with the role of response-produced stimuli in generalization and discrimination phenomena, verbal rote learning, and other complex behaviors.[2] Of these treatments, Dollard and Miller's mechanisms or paradigms of acquired equivalence and acquired distinctiveness of cues seem to provide the most explicit and general stimulus-response scheme. However, because of their interest in applying these paradigms to social and neurotic behavior, they gave relatively less attention to explicit assimilation of the mechanisms within the network of laboratory-established principles of animal and human learning.

Accordingly, it appears desirable to extend the Dollard and Miller analysis in three directions: (*a*) to formulate a more general dimensional model—within which acquired equivalence and distinctiveness appear as special cases—of the interaction of external cues, response-produced cues, and instrumental responses; (*b*) by means of this dimensional model, taken in combination with general principles of response acquisition and retention, to derive illustrative relationships between the acquisition of cue-producing responses and subsequent learning of instrumental responses; and (*c*) within the framework of the dimensional scheme and learning principles, to reinterpret a number of special learning and thinking phenomena such as rehearsal, familiarity and meaningfulness of stimuli, and concept formation.[3] The first two of these tasks are essayed in the present paper; the third will be treated in a later publication.

DIMENSIONAL ANALYSIS

Elements in the mechanisms suggested by Dollard and Miller which provide the basis for a more general dimensional analysis of the functional significance of response-produced stimuli for instrumental behavior are external stimuli, cue-producing responses and resultant cues, and emotional or instrumental responses. In line with these authors, cue-producing responses will be defined functionally rather than topographically (3, p. 99). Verbal mediating responses and emotional or nonverbal instrumental responses were of greatest concern to Dollard and Miller and will be emphasized in the present context. However, manipulative and other nonverbal responses (1, 10) can presumably function as mediating processes in both human and infrahuman behavior.

---

[1] R. M. Gagné, I. L. (Rossman) Gerjuoy, R. B. Holton, and N. E. Miller (personal communications) have contributed suggestions which, while not specifically acknowledged, have been most useful. In addition, the author is greatly indebted to M. E. W. Goss, W. E. Jeffrey, D. S. Palermo, and B. J. Underwood for critical readings of parts or all of various versions of the manuscript.

[2] Since this article is not intended as an exhaustive survey of relevant literature, only selected studies and theoretical papers have been noted. Other pertinent materials have been noted in (1, 7, 11, 12, 17).

[3] A dimensional model is here considered as the outcome of an analysis of particular behavioral phenomena into stimulus and response variables and possible interrelationships among those dimensions.

Dollard and Miller's Mechanisms

Dollard and Miller have defined or specified the functional properties of acquired equivalence and acquired distinctiveness as follows:

> *Acquired equivalence.* "Attaching the same cue-producing response to two distinctive stimulus objects gives them a certain learned equivalence increasing the extent to which instrumental and emotional responses will generalize from one to the other" (3, p. 101).
>
> *Acquired distinctiveness.* "Attaching distinctive cue-producing responses to similar stimulus objects tends to increase their distinctiveness" (3, pp. 101–102).

These mechanisms have been diagrammed in Fig. 1 by the use of bright (S_B), medium (S_M), and dim (S_D) intensities of light from a single source as external stimuli, two verbal responses (R_{jer}, R_{nig}) each producing cues (S_{jer}, S_{nig}), and two instrumental responses of moving a lever attached to a universal joint to the right (R_R) or left (R_L).[4] The solid lines of the top diagram represent an acquired-equivalence condition involving two physically dissimilar lights, S_B and S_D, each of which evokes the sequence $R_{jer} \sim S_{jer}$. Because of the common verbal cue, S_{jer}, the resultant stimulus patterns, $S_B + S_{jer}$ and $S_D + S_{jer}$, are assumed to be more similar to each other than S_B and S_D alone. Therefore, as indicated by the dashed lines, conditioning $S_B + S_{jer}$ to R_L should result in greater probability that S_D, through the presence of S_{jer} in the pattern $S_D + S_{jer}$, would evoke R_L, than would have been the case had S_B alone been conditioned to R_L with a subsequent test for generalization of R_L to S_D.

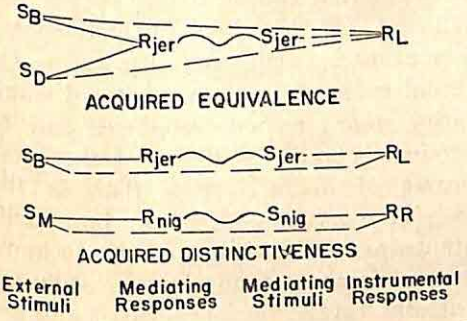


FIG. 1. Acquired-equivalence and acquired-distinctiveness relationships as postulated by Dollard and Miller.

An acquired-distinctiveness relationship is depicted in the bottom diagram. Two similar stimuli, S_B and S_M, lead to different and dissimilar response-produced cues, S_{jer} and S_{nig}, respectively. It is assumed that the resultant stimulus patterns, $S_B + S_{jer}$ and $S_M + S_{nig}$, are more dissimilar than S_B and S_M alone. Accordingly, less generalization of R_L from $S_B + S_{jer}$ to $S_M + S_{nig}$ and of R_R from $S_M + S_{nig}$ to $S_B + S_{jer}$ would be anticipated than under the condition of S_B and S_M alone as differential cues for the two instrumental responses. Also, because of greater distinctiveness, more rapid acquisition of $S_B + S_{jer} -- R_L$ and $S_M + S_{nig} -- R_R$ (broken lines) is predicted than of $S_B -- R_L$ and $S_M -- R_R$.

The Dimensional Model

Expansion of the number of stimulus-response components and more explicit consideration of degree of similarity of external and mediating cues will here constitute the initial steps in the development of a more general model. The role of response probability will then be considered. After acquired equivalence and acquired distinctiveness have been redefined within the expanded model and other relationships suggested, the final step of the dimensional expansion will be the assimila-

---

[4] The prototypes of Fig. 1 and subsequent diagrams were originated by Birge (1) and further developed by Thompson (17). In this article, the pairs of terms, cue-producing response and mediating response, and response-produced cues (stimuli) and mediating cues (stimuli), are used interchangeably.

LIBRARY 13.8.70 J 162

tion or representation within the model of additional variables and relationships.

Expansion

Number of components. In Fig. 2 the number of stimulus-response components has been increased from three to four different intensities of white light (very bright, S_{VB}; bright, S_B; dim, S_D; and very dim, S_{VD}) and from two to four nonsense syllable responses (R_{jer}, R_{ner}, R_{neg}, R_{nig}) of differing formal similarity each producing corresponding cues (S_{jer}, S_{ner}, S_{neg}, S_{nig}). Acquired equivalence is represented in the top diagram if it is assumed that the patterns $S_{VB} + S_{jer}$, $S_B + S_{jer}$, $S_D + S_{jer}$, $S_{VB} + S_{jer}$, which result from the evocation of R_{jer} by each of the four external stimuli, are more similar than the four lights alone. In the bottom diagram, the stimulus patterns $S_{VB} + S_{jer}$, $S_B + S_{ner}$, $S_D + S_{neg}$, $S_{VD} + S_{nig}$ are conceived as less similar than the four intensities alone. Thus, the addition of mediating cues presumably provides acquired distinctiveness or dissimilarity among the more similar external cues.

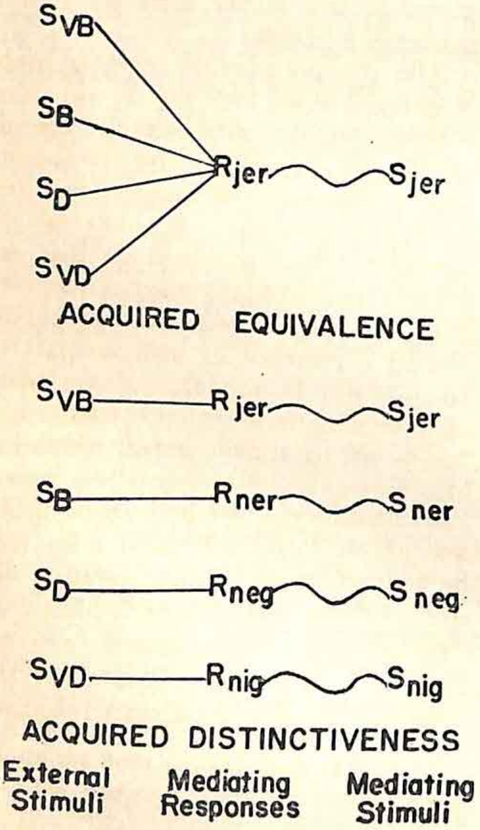


FIG. 2. Expansion of stimulus-response components of acquired-equivalence and acquired-distinctiveness paradigms.

With respect to acquired equivalence, greater similarity of the external cues should decrease the contribution of the common mediating component of the stimulus pattern, while less similarity of these cues should increase this contribution. The degree of similarity of both external and mediating cues is also relevant for acquired distinctiveness.[5] For example, high similarity among external cues will require greater dissimilarity among response-produced cues to produce stimulus patterns of the same degree of distinctiveness than in the case of external cues of an intermediate degree of similarity. Thus, in more general terms, similarity of stimulus patterns must be specified as some joint product of degree of similarity among sets of external and mediating cues.

Other combinations of external and mediating cues can be formed. Thus S_{VB} and S_B might both evoke $R_{jer} \sim S_{jer}$, while $R_{nig} \sim S_{nig}$ is elicited by

---

[5] Similarity can be expressed in physical terms as extent of differences along physical dimensions. Differences in numbers of common elements can also be used. Psychophysical definitions of similarity are functions of such differences, the experimenter's instructions, and subjects' responses. Also, although possibly subsumed under the instructions factor, mediating stimuli probably play a role in such determinations. For the analytical purpose of the present paper, however, it is assumed that the relative similarity of external and of mediating cues can, in principle, be specified separately with a minimum of involvement of other mediating cues.

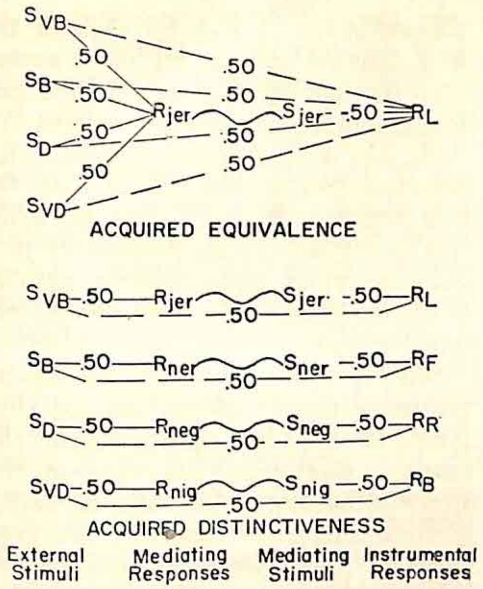


FIG. 3. Acquired equivalence and acquired distinctiveness including response probabilities and instrumental responses.

both S_D and S_{VD}. The equivalence of the S_{VB} and S_B pair and of the S_D and S_{VD} pair would then be increased, with a concomitant increase in the distinctiveness of the stimuli of the former pair with respect to those of the latter pair.

Instrumental responses and response probabilities. Movements of a lever to the left (R_L), forward (R_F), right (R_R), or backward (R_B) have been introduced as instrumental responses (Fig. 3). Also, *assumed* probabilities that external or mediating stimuli will evoke the responses have been added. These probabilities have been specified by the numbers in the breaks of the lines which connect various stimuli and responses. Specific assumptions concerning the probabilities that stimuli will evoke responses are: (*a*) the probability that a given stimulus will evoke a response varies from .00 to 1.00 (where no lines connect stimuli and responses, the probability of response occurrence to those stimuli is taken as .00); (*b*) external and/or mediating stimuli can elicit one, two, or more mediating or instrumental responses subject to the restriction that the separate probabilities summate to a maximum of 1.00; and (*c*) the sets of probabilities that external cues will arouse two or more mediating and/or instrumental responses are independent of the probabilities that response-produced cues will elicit one or more instrumental responses. For expository purposes these probabilities are taken as resultant probabilities. That is, they are considered the outcome of the absolute and relative strengths of the habits for competing responses (8) or of trial-by-trial variations in samples of elements (4). It is also assumed that each mediating response will evoke only one mediating stimulus, this sequence being given a probability of 1.00.

The incorporation of response probabilities within the model will ultimately require more specific consideration of procedures for combining separate probabilities that external and mediating cues elicit instrumental responses.[6] With probabilities that external cues evoke instrumental responses held constant, however, the probabilities of occurrence of instrumental responses will be products of the separate probabilities of external cues arousing mediating stimuli and of mediating stimuli evoking instrumental responses.

Furthermore, acquired equivalence and acquired distinctiveness must now be treated in probability terms. Because of the assumed probabilities of .50 for associations between external cues and mediating responses in the diagrams of Fig. 3, acquired equivalence or acquired distinctiveness will operate on only one-half of the occurrences of external stimuli. Of course, the probabilities of .50 that external and mediating stimuli elicit instrumental responses introduce additional contingencies.

---

[6] Hull's (8) procedure for habit summation represents one attempt to deal with this general problem. Or, the problem might reduce to a solution such as a larger "*N*" within Estes and Burke's (4) statistical conceptic of learning.

Any values from .00 to 1.00 could have been substituted in Fig. 3 for the illustrative probabilities of .50. With stimulus-response associations thus conceived, acquired equivalence and acquired distinctiveness appear as special cases or extremes of phenomena existing in degree. Moreover, in order for either mechanism to play a significant role in the acquisition and evocation of instrumental responses, the probabilities that given external stimuli will arouse cue-producing responses should approach 1.00.

Recognition should also be given to the possibility that low probabilities for cue-producing responses of acquired equivalence or distinctiveness and high probabilities of occurrences of other verbal responses might retard the acquisition of discriminative or generalized instrumental responses. In the acquired equivalence diagram of Fig. 3, for example, the arousal of R_{nig} by S_{VD} instead of R_{jer} should decrease generalization of R_L from S_{VB} to S_{VD}. Or, in the bottom diagram, because of the similarity of S_{neg} and S_{nig}, arousal of R_{neg} by S_{VD} might increase the similarity of S_{VD} and S_D and thus slow down the acquisition of the discriminative instrumental responses, R_R and R_B, to S_D and S_{VD}, respectively.

Additional Variables and Relationships

The significance of stimulus complexity, relationships among cue-producing responses, and "warm-up" phenomena for acquired equivalence and distinctiveness has received little attention. As a consequence, the steps which follow are preliminary efforts to provide more explicit representation of these variables and conditions.

Stimulus complexity. The stimulus-response conditions of learning tasks are rarely as simple as those in Fig. 1, 2, and 3. External cues can differ multidimensionally; each one may be some unique combination of values along two or more dimensions. Spatial and/or temporal patterns are a second source of complexity. A third source of complexity is represented in the two sets of differential external stimuli of many instrumental response tasks. One set is the lights or other stimuli which, under appropriate instructions, arouse cue-producing and/or instrumental responses. Many motor tasks also involve sets of levers, knobs, buttons, etc. whose selection and manipulation constitute instrumental responses. Thus, each of the actual differential cues of many instrumental tasks might be considered a two-component pattern made up of one of the manipulanda for the instrumental responses and a paired, spatially separated "signalling" cue such as one of the lights.[7]

These variations, alone or in combination, raise the possibility that each complex external cue of verbal or motor tasks will elicit several verbal responses. Some verbal responses of varying probabilities of occurrence will probably be unique to a given complex cue, while others might be aroused by two or more of the complex cues.

Figure 4 schematizes relationships which might be obtained for stimulus complexity based on multidimensional variation. The external stimuli are diagrammed as differing with respect to intensity, wave length, and simple form. Probabilities with which each stimulus arouses verbal responses for differences

---

[7] If receptor-orienting responses (**16**, **18**) must be learned to assure differential stimulation, an additional set of response-produced cues is introduced.

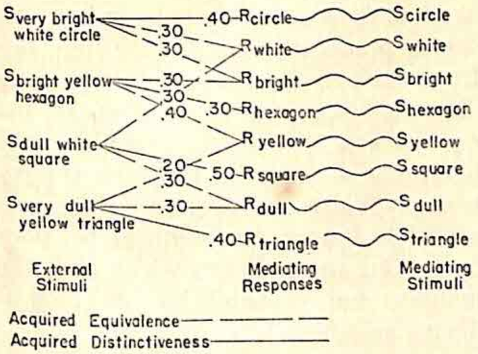


FIG. 4. Acquired-equivalence and acquired-distinctiveness relationships for complex stimuli.

along each of these dimensions have also been introduced. High probabilities for form responses (R_{circle}, $R_{hexagon}$, etc.) would provide acquired distinctiveness. On the other hand, if the two bright and the two dull, or the two white and the two yellow, evoked their common responses with high probabilities, some degree of acquired equivalence would obtain. Because of the low response probabilities which have been assumed, however, neither mechanism would be consistently operative.

Complexity of mediating cues must also be recognized. Verbal mediating cues can differ multidimensionally with respect to duration and amplitude. Or, such cues might be single letters or numbers or increasingly complex combinations of such elements as nonsense syllables, words, or identifying numbers. Furthermore, response-produced cues can presumably elicit other cue-producing responses. Therefore, varying degrees of complexity might arise from patterns of cues produced by sequences of cue-producing responses.

Relationships between mediating responses. Associations between response-produced cues and cue-producing responses might have other effects. Thus, in Fig. 3 if S_{nig} came to elicit R_{neg}, the acquired distinctiveness of S_{VD} and S_D would be reduced. Although not diagrammed in Fig. 3, occurrences of parallel sequences of different cue-producing responses such as S_{VD} leading to $R_{nig} \sim S_{nig}$ —— $R_{sil} \sim S_{sil}$ —— $R_{nok} \sim S_{nok}$ and S_D arousing $R_{neg} \sim S_{neg}$ —— $R_{guz} \sim S_{guz}$ —— $R_{das} \sim S_{das}$, by adding further distinctive response-produced cues to the stimulus patterns, should produce greater dissimilarity. Also possible are situations in which dissimilar external cues elicit sequences of verbal responses that produce increasingly similar verbal stimuli and thereby enhance equivalence. Other relationships between cue-producing responses have been treated in detail by Cofer and Foley (2).

Warm-up, etc. It has been observed that pretask experiences which are relatively nonspecific to the stimulus-response relationships of the tasks sometimes influence subsequent acquisition or retention of such relationships. The apparent effects of these pretask or warm-up experiences have been explained by hypothesizing *learning-how-to-learn* and *learning set* processes. In turn, these hypothesized processes have been reduced to postural adjustments, receptor-orienting and exposure responses, etc. (9).

To the degree that pretask experiences lead to the development of postural adjustments, and to receptor-orienting and exposure responses, the proportions of learning or retention trials on which external stimuli are "paid attention to" or received and/or the duration of reception on each trial should increase. The more frequent and longer the reception of stimuli the greater the number of cue-producing responses which would be elicited on each trial. Furthermore, higher probabilities of the occurrence of a given response on a single trial as well as increments in the number of mediating cue and cue-producing response relationships would be anticipated. Thus, warm-up could be conceptualized within the dimensional model as increased numbers and probabilities of occurrence of cue-producing responses on each trial. In addition, for some tasks the acquisition of receptor-orienting responses would heighten the probability that critical differential cues would be received. Such responses would have the effect of increasing the dissimilarity of external cues (16, 18). Under some conditions, however, enhanced similarity might result.

The number and probabilities of occurrence of cue-producing responses might also be altered by experiences which are more specific to the external stimuli. For example, it has been observed that experiences in seeing or seeing and discriminating stimuli influence subsequent discriminative motor learning (7). The strengthening of responses of looking for the specific parts or part-characteristics which differentiate complex stimuli might also be relevant (14). Finally, experiences or instructions which alter the probabilities of occurrence of responses to the dimensions along which sets of external stimuli are the same or different should be considered.

Summary of Dimensional Scheme

The interactions of cue-producing and instrumental responses in complex behaviors can be dimensionalized in terms of the *number, similarity, probabilities of occurrence,* and *complexity* of the following factors: (*a*) external stimulus events, (*b*) cue-producing responses and response-produced cues occurring in the presence of given external stimuli, (*c*) cue-producing responses and resultant cues arising in relation to response-produced stimuli alone or in combination with other response-produced stimuli and/or external cues, and (*d*) instrumental or emotional responses as functions of external and/or response-produced stimuli. Within this framework Dollard and Miller's paradigms of acquired distinctiveness and equivalence appear as particular combinations of stimulus-response components. Warm-up experiences are presumed to influence subsequent behavior by leading to conditions which alter the number and probabilities of cue-producing responses.

DIMENSIONAL ANALYSIS AND LEARNING PRINCIPLES

The presence of acquired distinctiveness, acquired equivalence, or other verbal-motor response relationships depends on particular combinations of probabilities of occurrence of cue-producing and/or instrumental responses. These combinations can be expected to vary as functions of classes of conventional learning variables (e.g., parameters of reinforcement of component stimulus-response units). Therefore, the usefulness of the dimensional scheme, when supplemented by general learning principles (13), will be illustrated by presenting quasi-quantitative, semiformal derivations of relationships between the conditions under which cue-producing responses are acquired and subsequent learning of instrumental responses. Specifically, model derivations will be outlined for direction and amount of transfer to discriminative motor learning as a function of number of discriminative verbal learning trials, and as influenced by the similarity of external cues common to both verbal and motor tasks. Possible relationships between motor learning and selected additional variables will then be noted without, however, detailed development of specific hypotheses.

Model Derivations

Number of verbal learning trials. On the assumption that dissimilar verbal

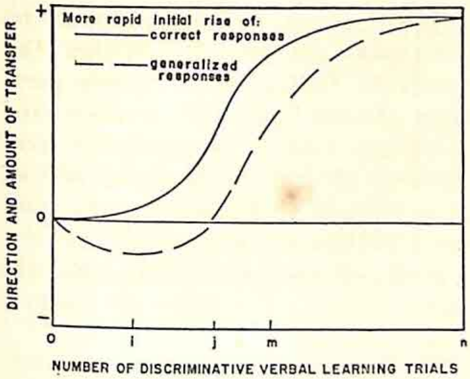


FIG. 5. Predicted general forms of the relationship between direction and amount of transfer and number of verbal learning trials.

responses to similar external stimuli increase differences in stimulus patterns, increased probabilities of verbal responses should occasion variation in the rate of learning of discriminative instrumental responses to the same stimuli. In order to specify direction and amount of transfer, however, possible effects of generalized responses must be considered. Because of primary stimulus generalization, each external cue will elicit not only its paired ("correct," dissimilar, or discriminative) response but also generalized responses (errors of generalization, "confusions," intralist intrusions). Enhanced similarity arising from acquired equivalence attributable to generalized responses will tend to offset increased dissimilarity of external and verbal cue patterns based on the acquisition of dissimilar labels.

Gibson (6) has hypothesized that while frequency of correct responses increases monotonically as a function of trials, the curve for generalized responses increases to a maximum and then decreases. If both generalized and correct responses increase initially, however, the interfering effects of the generalized responses should preclude marked positive transfer to discriminative motor tasks until sufficient trials have been given for these responses to begin to decline in frequency. If generalized responses increase more rapidly than correct responses, small numbers of verbal learning trials might lead to some negative transfer. Relative rates of rise of correct and generalized responses (and, by implication, the forms of transfer curves) are considered functions of degree of similarity of external cues. It has been assumed that the curves of Fig. 5 approximate those which might be obtained with external stimuli of an intermediate degree of similarity. Contingent on the relative rates of increase of correct and generalized responses, hypothesized directions and amounts of transfer to discriminative motor learning have been plotted as functions of number of discriminative verbal learning trials. A more rapid initial rise of correct responses should result in the solid-line relationship. The broken-line curve might be obtained with steeper initial increases in generalized responses.

Although consistent with the notion of a slow initial rise, neither the Gagné and Baker (5) nor the Rossman and Goss (14) findings are based on sufficient points to decide between these alternative curves. Because only high

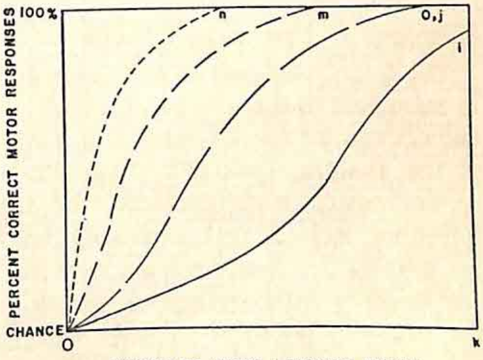


FIG. 6. Predicted family of curves for percentages of correct motor choices as a function of number of motor learning trials with amount of prior verbal training as the parameter.

levels of verbal response strength or one degree of verbal learning was employed, the results of other studies cannot be used for such a decision.

Proceeding on the basis of the broken-line or negative-transfer curve, number of verbal learning trials was used to generate a hypothetical family of curves relating percentages of correct motor discriminations to number of motor learning trials (Fig. 6). The slow initial rises for the curves with *i*, 0, *j*, and *m* discriminative verbal learning trials as parameters also stem from the assumption that generalized responses interfere with facilitation attributable to different verbal responses to each stimulus. Thus the greater the number of generalized responses, the greater and more prolonged the interference with acquisition of motor discriminations. Rossman and Goss (14) and Smith and Goss (15) obtained curves with slow initial rises in correct motor discrimination measures. While Gagné and Baker's (5) curves for response times are negatively accelerated, the plots of errors for the 0-, 8-, and 16-trial verbal learning conditions could be the initial segments of **S**-shaped curves. The curve for correct responses of McAllister's (11) group 2 exhibits relative flatness over the first three trials followed by a negatively accelerated increase.

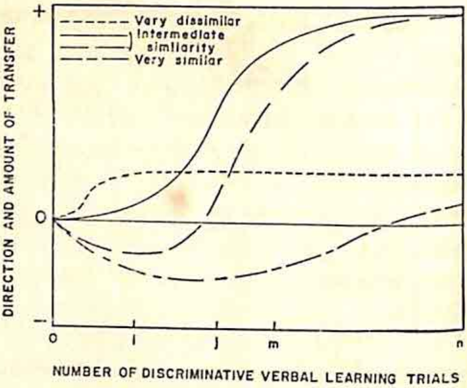


FIG. 7. Predicted family of curves for direction and amount of transfer to a motor task as a function of the number of discriminative verbal learning trials with similarity of stimuli as the parameter.

Similarity of external stimuli. Disregarding prior verbal learning, the difficulty of discriminative motor learning should, through changes in the relative frequencies of occurrence of correct and generalized verbal responses, vary with similarity of external cues. Specifically, increasing similarity of external cues should lead to lower rates of rise and to lower final asymptotes for correct verbal response curves. Generalized verbal responses would be expected to rise more rapidly to higher levels and then to decrease from the maxima less rapidly as a function of increasing similarity.

Figure 7 represents hypothesized relationships between transfer and number of verbal learning trials for sets of external stimuli which are assumed to be very dissimilar, of an intermediate degree of similarity, or very similar. Although plotted against the same zero (0) transfer line, a different control group would be necessary for each degree of similarity. Negligible occurrence and strengthening of mediating responses during the course of instrumental learning have been assumed. Because dissimilarity of external cues should result in relatively infrequent generalized verbal responses, it is probable, as suggested by the dotted-line curve, that only facilitation with little or no positive acceleration would be observed. The solid- and broken-line curves which duplicate those of Fig. 5 are alternative representations of transfer as a function of verbal learning experiences with stimuli of some intermediate degree of similarity.

Very similar external stimuli should occasion a rapid initial rise in generalized responses to a high level and a relatively small and slow subsequent decrement. Therefore, the remaining curve depicts the possibility that high frequencies of generalized responses might lead to negative transfer effects which would persist despite many verbal learning trials. Moreover, the relatively small decrements in generalized responses might lead to a fairly low asymptote for any eventual positive transfer.

Relationships Involving Other Variables

Neither space nor available data justify the continuation of relatively explicit and complete derivations on a variable-by-variable basis. It would seem desirable, however, to illustrate further the range of relationships between acquisition of cue-producing and instrumental responses which can be developed by means of the dimensional model and learning principles. Therefore, relationships involving four arbitrarily selected variables or conditions of learning prior to acquisition of instrumental responses to external cues will be treated. These variables or conditions are: (*a*) similarity of response-produced cues, (*b*) inhibition of cue-producing responses, (*c*) relationships between response-produced cues and instrumental responses, and (*d*) changes in the similarity of the external stimuli of the tasks of first learning cue-producing responses and then acquiring instrumental responses.

Similarity of verbal response-produced cues. Because similarity of cue patterns was conceived as determined by similarity of both external and mediating stimuli, changes in similarity of the latter cues should influence acquisition of instrumental responses. For example, with dissimilar external cues held constant it would be predicted that an increase in the similarity of response-produced cues would shift the hypothesized relationship between transfer and number of discriminative verbal learning trials from the dotted-line curve of Fig. 7 toward the alternative forms for external cues of intermediate similarity. Eventually, the curve for very similar external cues might be approximated.

Inhibition of cue-producing responses. Rossman and Goss (14) and Murdock (12) have reported that conditions (electric shock, failure, acquisition of competing responses) intervening between verbal and motor learning which lead to inhibition of verbal responses may decrease both discriminative motor learning and generalization of instrumental responses. Accordingly, it appears likely that instrumental-response learning will be influenced by conditions which inhibit cue-producing responses. The specific consequences of such inhibition will, of course, depend on the functional role of the cue-producing responses in the learning of instrumental responses.

Relationships between response-produced cues and instrumental responses. Direction and amount of transfer to motor learning tasks should vary with conditions of acquisition and resultant strengths of associations of mediating cues and instrumental responses. Thus, given dissimilar verbal responses, each elicited with high probability by different external cues of intermediate similarity, facilitation of learning of discriminative instrumental responses should increase with amount of prior experience with a task requiring different instrumental responses to different response-produced cues. McAllister's (11) *relevant* S-R condition apparently assumes relationships between mediating stimuli and discriminative instrumental responses. If types of *relevant* S-R relationships are interpreted as also involving different *degrees* of strength of mediating cue-instrumental response relationships, her results are consistent with this notion.

Similarity of external stimuli of cue-producing-response and instrumental-response learning. The relationships treated up to this point rest on the condition that the same external stimuli, mediating responses and stimuli, and instrumental responses are present during the strengthening of associations between these components. However, intertask changes in similarity and/or number of any one or any combination of stimulus-response elements could be expected to affect the acquisition of discriminative instrumental responses. Discriminative verbal response to external stimuli of high or intermediate similarity may have been learned to a high level of mastery. As similarity between those stimuli and external cues for a subsequent discriminative motor task decreases, a monotonic decrease in dissimilar verbal responses to each external cue and an increase and subsequent decrease in generalized responses should be observed. Therefore, amount of positive transfer should decrease and, contingent on changes in the relative frequencies of discriminative and generalized responses, some negative transfer might be obtained before reaching a zero transfer point.

SUMMARY

Dollard and Miller's mechanisms of the acquired equivalence and acquired distinctiveness of cues served as bases for a more general dimensional analysis of the interaction of cue-producing and instrumental responses. With external cues, cue-producing responses and resultant cues, and instrumental responses as stimulus-response components, the model provided for the effects of response probabilities, stimulus complexity, relationships between cue-producing responses, and conditions such as warm-up.

The dimensional model and general principles of response acquisition and retention were linked by means of response probabilities. It was then possible to outline model derivations of the influence of number of discriminative verbal learning trials and of similarity of external cues on rate of acquisition of discriminative instrumental responses. Brief attention was also given to the learning of instrumental responses as a function of: (*a*) similarity of response-produced cues, (*b*) inhibition of cue-producing responses, (*c*) relationships between response-produced cues and instrumental responses, and (*d*) similarity of external cues for cue-producing and subsequent instrumental responses.

REFERENCES

1. BIRGE, J. S. The role of verbal responses in transfer. Unpublished Ph.D. dissertation, Yale Univer., 1941.
2. COFER, C. N., & FOLEY, J. P. Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.*, 1942, **49**, 513–540.
3. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
4. ESTES, W. K., & BURKE, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.
5. GAGNÉ, R. M., & BAKER, K. E. Stimulus pre-differentiation as a factor in the transfer of training. *J. exp. Psychol.*, 1950, **40**, 439–451.
6. GIBSON, ELEANOR J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
7. GOSS, A. E. Transfer as a function of type and amount of preliminary experience with task stimuli. *J. exp. Psychol.*, 1953, **46**, 419–428.
8. HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
9. IRION, A. L. The relation of 'set' to retention. *Psychol. Rev.*, 1948, **55**, 336–341.
10. LAWRENCE, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39**, 770–784.

11. McALLISTER, DOROTHY E. The effects of various kinds of relevant verbal pretraining on subsequent motor performance. *J. exp. Psychol.,* 1953, **46**, 329–336.
12. MURDOCK, B. B., JR. The effects of failure and retroactive inhibition on mediated generalization. *J. exp. Psychol.,* 1952, **44**, 156–164.
13. OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.
14. ROSSMAN, I. L., & GOSS, A. E. The acquired distinctiveness of cues: the role of discriminative verbal responses in facilitating the acquisition of discriminative motor responses. *J. exp. Psychol.,* 1951, **42**, 173–182.
15. SMITH, S. L., & GOSS, A. E. The role of the acquired distinctiveness of cues in the acquisition of a motor skill in children. *J. genet. Psychol.,* in press.
16. SPENCE, K. W. Continuous versus noncontinuous interpretations of discrimination learning. *Psychol. Rev.,* 1940, **47**, 271–288.
17. THOMPSON, L. The role of verbalization in learning from demonstrations. Unpublished Ph.D. dissertation, Yale Univer., 1944.
18. WYCKOFF, L. B., JR. The role of observing responses in discrimination behavior. *Psychol. Rev.,* 1952, **59**, 437–442.

(Received March 15, 1954)

Psychological Review
Vol. 62, No. 1, 1955

PERCEPTUAL LEARNING: DIFFERENTIATION OR ENRICHMENT?[1]

JAMES J. GIBSON AND ELEANOR J. GIBSON

Cornell University

The term "perceptual learning" means different things to different psychologists. To some it implies that human perception is, in large part, learned—that we learn to see depth, for instance, or form, or meaningful objects. In that case the theoretical issue involved is *how much* of perception is learned, and the corresponding controversy is that of nativism or empiricism. To others the term implies that human learning is in whole or part a matter of perception—that learning depends on comprehension, expectation, or insight, and that the learning process is to be found in a central process of cognition rather than in a motor process of performance. In this second case, the theoretical issue involved is whether or not one has to study a man's perceptions before one can understand his behavior, and the controversy is one of long standing which began with old-fashioned behaviorism.

These two sets of implications are by no means the same, and the two problems should be separated. The problem of the role of learning in perception has to do with perception and the effect of past experience or practice on it. The problem of the role of perception in learning has to do with behavior and the question of whether we can learn to do something by perceiving, or whether we can only learn by doing it. The questions, then, are these: (*a*) In what sense do we learn to perceive? (*b*) In what sense can we learn by perceiving? Both questions are important for the practical problems of education and training, but this paper will be concerned with the former.

IN WHAT SENSE DO WE LEARN TO PERCEIVE?

This question has roots in philosophy and was debated long before experimental psychology came of age. Does all knowledge (information is the contemporary term) come through the sense organs or is some knowledge contributed by the mind itself? Inasmuch as sensory psychology has been unable to explain how as much information about the world as we manifestly do obtain is transmitted by the receptors, some theory is required for this unexplained surplus. There has been a variety of such theories ever since the days of John Locke. An early notion was that the surplus is contributed by the rational faculty (rationalism). Another was that it comes from innate ideas (nativism). In modern times there have been few adherents to these positions. The most popular theory over the years has been that this supplement to the sensations is the result of learning, and that it comes from past experience. A contemporary formula for this explanation is that the brain stores information—possibly in the form of traces or memory images, but conceivably as attitudes, or mental sets, or general ideas, or concepts. This approach has been called empiricism. It preserves the dictum that all knowledge comes from experience by assuming that past experience somehow gets *mixed with* present experience. It assumes, in other words, that experience *accumulates*, that traces of the past somehow exist in our perception of the present. One of its high-water marks was Helmholtz's theory of unconscious inference, which supposes that we learn to see depth by interpreting the clues furnished by the depthless sensations of color. Another was Titchener's context theory of meaning, which asserts that we learn to perceive objects when a core of sensations acquires by association a context of memory images.

---

[1] This paper is a revision, with added experimental material, of one given in May 1953 at a symposium on the psychology of learning basic to problems of military training (8) conducted by the Panel on Training and Training Devices of the Research and Development Board, Washington, D. C.

Over a generation ago this whole line of thought was challenged by what seemed to be a different explanation for the discrepancy between the sensory input and the finished percept—the theory of sensory organization. The gestalt theorists made destructive criticisms of the notion of *acquired* linkages among sensory elements and their traces. Instead they asserted that the linkages were *intrinsic*, or that they arose *spontaneously*, taking visual forms as their best example. Perception and knowledge, they said, were or came to be *structured*.

The theory of sensory organization or cognitive structure, although it generated a quantity of experimentation along new lines, has not after 30 years overthrown the theory of association. In this country the old line of empiricist thinking has begun to recover from the critical attack, and there are signs of a revival. Brunswik (2, pp. 23 ff.) has followed from the start the line laid down by Helmholtz. Ames and Cantril and their followers have announced what might be called a neoempiricist revelation (3, 11, 14). Other psychologists are striving for a theoretical synthesis which will include the lessons of gestalt theory but retain the notion that perception is learned. Tolman, Bartlett, and Woodworth began the trend. Leeper took a hand in it at an early date (15). The effort to reconcile the principle of sensory organization with the principle of determination by past experience has recently been strenuously pursued by Bruner (1) and by Postman (16). Hilgard seems to accept both a process of organization governed by relational structure and a process of association governed by the classical laws (10). Hebb has recently made a systematic full-scale attempt to combine the best of gestalt theory and of learning theory at the physiological level (9). What all these theorists seem to us to be saying is that the organization process and the learning process are not inconsistent after all, that both explanations are valid in their way, and that there is no value in continuing the old argument over whether learning is really organization or organization is really learning. The experiments on this issue (beginning with the Gottschaldt experiment) were inconclusive, and the controversy itself was inconclusive. Hence, they argue, the best solution is to agree with both sides.

It seems to us that all extant theories of the perceptual process, including those based on association, those based on organization, and those based on a mixture of the two (including attitudes, habits, assumptions, hypotheses, expectation, images, contexts, or inferences) have at least this feature in common: they take for granted a discrepancy between the sensory input and the finished percept and they aim to explain the difference. They assume that somehow we get more information about the environment than can be transmitted through the receptor system. In other words, they accept the distinction between sensation and perception. The development of perception must then necessarily be one of supplementing or interpreting or organizing.

Let us consider the possibility of rejecting this assumption altogether. Let us assume tentatively that the stimulus input contains within it everything that the percept has. What if the flux of stimulation at receptors *does* yield all the information anyone needs about the environment? Perhaps all knowledge comes through the senses in an even simpler way than John Locke was able to conceive—by way of variations, shadings, and subtleties of energy which are properly to be called stimuli.

THE ENRICHMENT THEORY VERSUS THE SPECIFICITY THEORY

The entertaining of this hypothesis faces us with two theories of perceptual learning which are clear rather than vague alternatives. It cuts across the schools and theories, and presents us with an issue. Is perception a creative process or is it a discriminative process? Is learning a matter of enriching previously meagre sensations or is it a matter of differentiating previously vague impressions? On the first alternative we might learn to perceive in this sense: that percepts change over time by acquiring progressively more memory images, and that a context of memories accrues by association to a sensory core. The theorist can substitute attitudes or inferences or assumptions for images in the above Titchenerian proposition, but perhaps all this does is to make the theory less neat while making the terminology more fashionable. In any case perception is progressively in *decreasing correspondence with stimulation.* The latter point is notable. Perceptual learning, thus conceived, necessarily consists of experience becoming more imaginary, more assumptive, or more inferential. The dependence of perception on learning seems to be contradictory to the principle of the dependence of perception on stimulation.

On the second alternative we learn to perceive in this sense: that percepts change over time by progressive elaboration of qualities, features, and dimensions of variation; that perceptual experience even at the outset consists of a world, not of sensation, and that the world gets more and more properties as the objects in it get more distinctive; finally, that the phenomenal properties and the phenomenal objects correspond to physical properties and physical objects in the environment *whenever learning is successful.* In this theory perception gets richer in differential responses, not in images. It is progressively in *greater* correspondence with stimulation, not in less. Instead of becoming more imaginary it becomes more discriminating. Perceptual learning, then, consists of responding to variables of physical stimulation not previously responded to. The notable point about this theory is that learning is always supposed to be a matter of improvement—of getting in closer touch with the environment. It consequently does not account for hallucination or delusions or, in fact, for any kind of maladjustment.

The latter kind of theory is certainly worth exploring. It is not novel, of course, to suggest that perceptual development is a matter of differentiation. As phenomenal description this was asserted by some of the gestalt psychologists, notably Koffka and Lewin. (Just how differentiation was related to organization, however, was not clear.) What *is* novel is to suggest that perceptual development is always a matter of the correspondence between stimulation and perception—that it is strictly governed by the relationships of the perceiver to his environment. The rule would be that, as the number of distinct percepts a man can have increases, so also the number of different physical objects to which they are specific increases. An example may clarify this rule. One man, let us say, can identify sherry, champagne, white wine, and red wine. He has four percepts in response to the total possible range of stimulation. Another man can identify a dozen types of sherry, each with many varieties, and numerous blends, and so on for the others. He has four thousand percepts in response to the range of stimulation. The crucial question to ask about this example of differentiated perception is its relation to stimulation.

Stimulus is a slippery term in psychology. Properly speaking stimulation is always energy at receptors, that is, proximal stimulation. An individual is surrounded by an array of energy and immersed in a flow of it. This sea of stimulation consists of variation and invariants, patterns and transformations, some of which we know how to isolate and control and others of which we do not. An experimenter chooses or constructs a sample of this energy when he performs a psychological experiment. But it is easy for him to forget this fact and to assume that a glass of wine is a stimulus when actually it is a complex of radiant and chemical energies which is the stimulus. When the psychologist refers to stimuli as cues, or clues, or carriers of information he is skipping lightly over the problem of how stimuli come to *function* as cues. Energies do not have cue properties unless and until the differences in energy have correspondingly different effects in perception. The total range of physical stimulation is very rich in complex variables and these are theoretically capable of becoming cues and constituting information. This is just where learning comes in.

All responses to stimulation, including perceptual responses, manifest some degree of specificity, and, inversely, some degree of nonspecificity. The gentleman who is discriminating about his wine shows a high specificity of perception, whereas the crude fellow who is not shows a low specificity. A whole class of chemically different fluids is equivalent for the latter individual; he can't tell the difference between claret, burgundy, and chianti; his perceptions are relatively undifferentiated. What has the first man learned that the second man has not? Associations? Memories? Attitudes? Inferences ? Has he learned to have perceptions instead of merely sensations? Perhaps, but a simpler statement might be made. The statement is that he has learned to taste and smell more of the qualities of wine, that is, he discriminates more of the variables of chemical stimulation. If he is a genuine connoisseur and not a fake, one combination of such variables can evoke a specific response of naming or identifying and another combination can evoke a different specific response. He can consistently apply nouns to the different fluids of a class and he can apply adjectives to the differences between the fluids.

The classical theory of perceptual learning, with its emphasis on subjective determination of perception in contrast to stimulus determination, gets its plausibility from experiments on errors in form perception, from the study of illusions and systematic distortions, and from the fact of individual differences in and social influences on perception. The learning process is assumed to have occurred in the past life of the experimental subject; it is seldom controlled by the experimenter. These are *not* learning experiments insofar as they do not control practice or take measures before and after training. True perceptual learning experiments are limited to those concerned with discrimination.

One source of evidence about discriminative learning comes from the study of the cues for verbal learning.

The analysis of these cues made by one of the authors in terms of stimulus generalization and differentiation (4) suggests the present line of thought. It has also led to a series of experiments concerned with what we call *identifying responses*. Motor reactions, verbal reactions, or percepts, we assume, are identifying responses if they are in specific correspondence with a set of objects or events. Code learning (13), aircraft recognition (7), and learning to name the faces of one's friends are all examples of an increasingly specific correspondence between the items of stimulation presented and the items of response recorded. As a given response gains univocality, the percept is reported to gain in the feeling of familiarity or recognition and to acquire meaning.

An Illustrative Experiment [2]

In order to provide a clear example of such learning, we studied the development of a single identifying response. The *S* was presented with a visual item consisting of a nonsense "scribble"; his recognition of it was tested when it was interspersed in a series of similar scribbles, and then the single showing and the multiple presentation were repeated until the item could be identified. We devised a set of 17 scribbles intended to be indistinguishable from the critical item on the first trial, and another set of 12 items intended to be distinguishable from the critical item on the first trial.

The items which had to be differentiated are shown in Fig. 1. The critical item, a four-coil scribble, is in the center and 16 other items are arranged outward from it. The eighteenth item (a reversal of the critical item) is not shown. It may be noted that there are three dimensions of variation from the critical item: (*a*) number of coils—three, four, or five, (*b*) horizontal compression or stretching, and (*c*) orientation or right-left reversal. The latter two kinds of variation were produced by photographic transformation. There are three degrees of coil frequency, three degrees of compression, and two types of orientation, which yields 18 items. Since one of these is the critical item, 17 remain for use in the experiment. The reader may observe that when these differences are verbally specified and the figures are displayed for immediate comparison, as in Fig. 1, they are clearly distinguishable. The *Ss* of the experiment, however, saw the items only in succession.

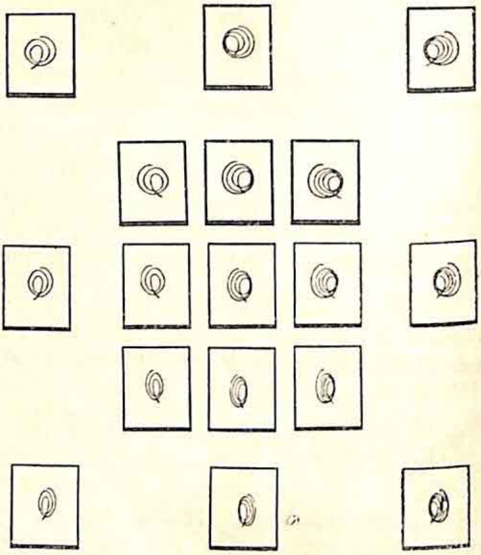


Fig. 1. Nonsense items differing in three dimensions of variation.

The 12 additional items presented on each recognition trial are shown in Fig. 2. Each differs from every other and from all of the

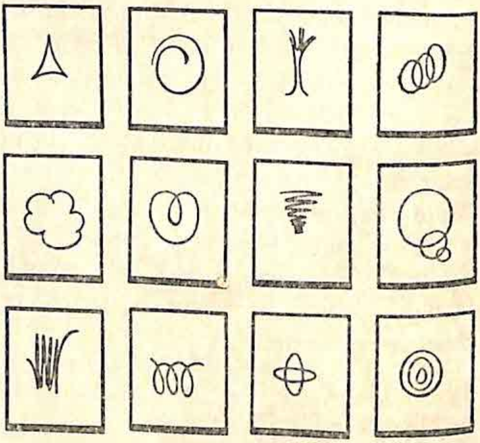


Fig. 2. Nonsense items differing in many dimensions of variation.

---

[2] This experiment was first reported at the meeting of the American Psychological Association in September 1950 in a paper read by Eleanor J. Gibson, and an abstract has been published (6).

set of 18. The differences from the scribbles were intended to be sufficient to make them appear different at the outset to *S*s with a normal amount of experience with drawn forms. The 30 items (12 plus 18) were printed photographically on stiff 2 in. × 4 in. cards with black borders, and made into a pack. The material available for any one learning trial consisted of the critical item plus a shuffled pack of cards among which were interspersed four replicas of the critical item.

The *S* was shown the critical item for about 5 sec. and told that some of the items in the pack would be exactly like the one shown. The series of 34 was then presented each with a 3-sec. exposure and *S* was asked to report which of them were the same figure. The identifying response recorded was any report such as "that's it" or "this is the one I saw before." The *S* was never told whether an identification was correct or incorrect. A record was kept not only of the identifying responses, but also of any spontaneous *descriptions* offered by *S*, which were later classified as *naming* responses and *qualifying* responses.

At the end of the first trial the critical figure was presented a second time and another shuffled pack was run through. The procedure of examining a figure and then trying to identify it when mixed with a series including figures of both great and little similarity was continued until *S* made only the four correct identifications in one trial. Three groups took part in the experiment: 12 adults, 10 older children (8½ to 11 years), and 10 younger children (6 to 8 years).

Results. In this experiment, learning is taken to be an increase in the specificity of an identifying response or, in other words, a decrease in the size of the class of items that will elicit the response. The data therefore consist of the number of items (out of a probable maximum of 17) reacted to as if they were the critical figure. As will be evident, this class of undifferentiated items was reduced as a result of repetition. The three groups of *S*s, however, began to learn at very different levels and learned at very different rates. The results are given in Table 1. For adults, the class of undifferentiated items at the outset was small (Mean = 3.0), and only a few trials were needed before this class was reduced to the critical item alone (Mean = 3.1). Two of these adults were able to make no other than correct identifying responses on the first trial. Both were psychologists who could have had previous acquaintance with nonsense figures. The learning task was so easy for this group that not much information about the learning process could be obtained. At the other extreme, however, the younger children "recognized" nearly all of the scribbles on the first trial (Mean = 13.4), which

TABLE 1

INCREASE IN SPECIFICITY OF AN IDENTIFYING RESPONSE FOR THREE AGE GROUPS

Variable	Adults ($N = 12$)	Older Children ($N = 10$)	Younger Children ($N = 10$)
Mean number of undifferentiated items on first trial	3.0	7.9	13.4
Mean number of trials required for completely specific response	3.1	4.7	6.7*
Percentage of erroneous recognitions for items differing in *one* quality	17	27	53
Percentage of erroneous recognitions for items differing in *two* qualities	2	7	35
Percentage of erroneous recognitions for items differing in *three* qualities	0.7	2	28

* Only two of the younger children achieved a completely specific identification. The mean number of undifferentiated items on the last trial was still 3.9.

is to say that the class of undifferentiated items was large. The number of trials needed to reduce this class to the correct item was so great that most of the *S*s could not be required to complete the experiment. Two out of 10 reached the criterion, but for the remainder the trials had to be stopped for reasons of fatigue. After an average of 6.7 trials the mean number of undifferentiated items was still 3.9. One child had so much difficulty with the task that *E* finally gave differential reinforcement by saying "right" or "wrong" after each presentation of a card. Although this procedure helped, wholly specific identifications were never achieved. The failures of the younger children to discriminate did not seem to be due merely to "inattention"; they understood that they were to select only the figures which were *exactly* the same as the critical figure.

For the older children (between 8½ and 11 years of age) the results were intermediate between these extremes. For them the particular task and the particular items were neither too hard nor too easy. The average number of undifferentiated items on the first trial was 7.9, and all children succeeded in reducing this to a single item after a mean of 4.7 trials.

Table 1 also indicates for each group an important fact about the unspecific responses: they tend to occur more often as the differences between the test item and the critical item become fewer. As Fig. 1 shows, a given scribble may differ in *one* quality or dimension (thickness, coil frequency, or orientation), or in *two* of these qualities, or in all *three* of them. Five of the scribbles differ in one feature, eight differ in two features, and four differ in three features. It will be recalled that the 12 additional forms shown in Fig. 2 differed from the critical item with respect to *more* than three features. Amount of difference can be usefully stated as number of differing qualities or, conversely, amount of sameness as the fewness of differing qualities.[3] The lower half of Table 1 gives the percentage of occurrence of false recognitions in the case of scribbles with one quality different, with two qualities different, and with three qualities different. These percentages are based on the number of times the items in question were presented during the whole series of trials. The "dissimilar" figures, which had many qualities different, yielded a zero percentage of false recognitions except for a few scattered instances among the younger children.

Discussion. The results show clearly that the kind of perceptual learning hypothesized has occurred in this experiment. A stimulus item starts out by being indistinguishable from a whole class of items in the stimulus universe tested, and ends by being distinguishable from all of them. The evidence for this assertion is that the specificity of *S*'s identifying response has increased. What has happened to produce this result?

The *S*s were encouraged to describe all the items of each series as they were presented, and a special effort was made to obtain and record these spontaneous verbal responses for seven of the older children. In general they tended to fall into two types, either naming responses or qualifying responses. Considering only the responses to the 17 scribbles, the record showed that the frequency of the latter type increased during the progress of learning. Examples of the former are nouns like *figure 6, curl, spiral, scroll*. Examples of the latter are adjectival phrases like *too thin,*

---

[3] Experiments on primary stimulus generalization have usually varied the magnitude of a *single* difference, not the number of differences, between the critical stimulus and the undifferentiated stimulus. However, our method of quantifying "amount of difference" has much to recommend it.

rounder, reversed. It is notable that the latter are responses not to the item as such but to the relation between it and the critical item. They are analogous to differential judgments in a psychophysical experiment. An adjective, in general, is a response which is specific not to an object but to a property of two or more objects. It is likely, then, that the development of a specific response to an item is correlated with the development of specific responses to the qualities, dimensions, or variables that relate it to other items. The implication is that, for a child to identify an object, he must be able to identify the differences between it and other objects, or at least that *when* he can identify an object he *also* can identify its properties.

The verbal reactions of the children to the 17 scribbles, both naming and qualifying, could be categorized by *E* as specific or nonspecific to the item in question. These judgments were necessarily subjective, but they were carried out with the usual precautions. Although a single adjective cannot be specific to a single item, a combination of adjectives can be. An example of a nonspecific reaction is "another curlicue," and of a specific reaction is "this one is thinner and rounder." The latter sort may be considered a spontaneously developing identifying reaction, not of the "that's it" type, it is true, but nevertheless fulfilling our definition. The mean number of such verbal reactions on the first trial was 7.7 out of 17, or 45 per cent. The mean number of such reactions on the last trial was 16.5, or 97 per cent. This suggests that, as a single identifying response becomes increasingly specific to one member of a group of similar items, verbal identifying responses also tend to become specific to the other members of the group. As the class of indistinguishable items which will elicit one response is diminished, the number of responses which can be made to the class increases.

OTHER EVIDENCE

Another source of experimental evidence about perceptual learning comes from psychophysics. Contrary to what might be expected, psychophysical experimenters over the years have shown a lively interest in perceptual learning, or at least in the bettering of perceptual judgments with practice. One of the authors has recently surveyed this neglected literature insofar as it concerns improvement of perception or increase in perceptual skills (5). There is a great quantity of evidence about progressive change in acuity, variability, and accuracy of perception, including both relative judgments and absolute judgments. It proves beyond a shadow of doubt that the notion of fixed thresholds for a certain set of innate sensory dimensions is oversimplified. Discrimination gets better with practice, both with and without knowledge of results. An example may be taken from the two-point threshold on the skin.

As long ago as 1858 it was discovered that there is a certain distance at which two points are felt double by a blindfolded subject that is characteristic of the area of the skin tested. At the same time, it was found that only a few hours of practice in this discrimination had the effect of reducing the distance to half of what it had been (17). Later experiments showed that the lowering of the threshold continued slowly for thousands of trials; for instance, it might go from 30 mm. to 5 mm. during four weeks of training. Moreover, the improved discrimination transferred to other untrained areas of the skin, transfer being nearly complete for the bilaterally symmetrical area. It was found that blind subjects had very much lower thresholds than seeing subjects even at the beginning of testing

(12, 18). The experimental improvement was largely lost after a period of disuse. It seemed to depend on confirmation or correction of the judgment, or, in the absence of that, on the development of a sort of scale from "close together" to "far apart" (5). It is clear that any theory of supposedly distinct sensations of oneness and twoness never had any support from these data. As one writer put it, the observer adopts different and finer *criteria* of doubleness. What might these criteria be? We suggest that the stimulation is complex, not simple, and that the observer continues to discover higher-order variables of stimulation in it. The percept becomes differentiated.

CONCLUSION

There is no evidence in all of this literature on perceptual learning, nor is there evidence in the experiment reported in the last section, to *require* the theory that an accurate percept is one which is enriched by past experience, whereas a less accurate percept is one *not* enriched by past experience. Repetition or practice is necessary for the development of the improved percept, but there is no proof that it incorporates memories. The notion that learned perception is less and less determined by external stimulation as learning progresses finds no support in these experiments. The observer sees and hears more, but this may be not because he imagines more, or infers more, or assumes more, but because he discriminates more. He is more sensitive to the variables of the stimulus array. Perhaps the ability to summon up memories is merely incidental to perceptual learning and the ability to differentiate stimuli is basic. Perhaps the dependence of perception on learning and the dependence of perception on stimulation are not contradictory principles after all.

This theoretical approach to perceptual learning, it must be admitted, has points of weakness as well as points of strength. It accounts for veridical perception, but it does not account for misperception. It says nothing about imagination or fantasy, or wishful thinking. It is not an obviously useful approach for the study of abnormal behavior or personality, if one is convinced that a man's perceptions are the clues to his motives. But if one is concerned instead with the practical question of whether training can affect favorably a man's perception of the world around him, a very productive field for theory and experiment is opened up.

REFERENCES

1. BRUNER, J. S. Personality dynamics and the process of perceiving. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: an approach to personality*. New York: Ronald, 1951. Pp. 121–147.
2. BRUNSWIK, E. The conceptual framework of psychology. *Int. Encycl. unif. Sci.*, 1952, 1, No. 10.
3. CANTRIL, H., AMES, A., JR., HASTORF, A. H., & ITTELSON, W. H. Psychology and scientific research. *Science*, 1949, **110**, 461–464, 491–497, 517–522.
4. GIBSON, ELEANOR J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, 47, 196–229.
5. GIBSON, ELEANOR J. Improvement in perceptual judgments as a function of controlled practice or training. *Psychol. Bull.*, 1953, 50, 401–431.
6. GIBSON, ELEANOR J., & GIBSON, J. J. The identifying response: a study of a neglected form of learning. *Amer. Psychologist*, 1950, 7, 276. (Abstract)
7. GIBSON, J. J. (Ed.) *Motion picture testing and research*. Washington, D. C.: Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.*, No. 7.)
8. GIBSON, J. J., & GIBSON, ELEANOR J. Perceptual learning in relation to training. In *Symposium on psychology of learning basic to military training problems*. Dept. of Defense, HR-HTD-201-1, 1953, 151–159.
9. HEBB, D. O. *The organization of behavior*. New York: Wiley, 1949.

10. HILGARD, E. R. The role of learning in perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: an approach to personality.* New York: Ronald, 1951. Pp. 95–120.

11. ITTELSON, W. H. The constancies in perceptual theory. *Psychol. Rev.,* 1951, 58, 285–294.

12. JASTROW, J. Psychological notes on Helen Keller. *Psychol. Rev.,* 1894, **1**, 356–362.

13. KELLER, F. S. Studies in International Morse Code. I. A new method of teaching code reception. *J. appl. Psychol.,* 1943, 27, 407–415.

14. KILPATRICK, F. P. *Human behavior from the transactional point of view.* Hanover, N. H.: Institute for Associated Research, 1952.

15. LEEPER, R. A study of a neglected portion of the field of learning—the development of sensory organization. *J. genet. Psychol.,* 1935, 46, 41–75.

16. POSTMAN, L. Toward a general theory of cognition. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads.* New York: Harper, 1951. Pp. 242–272.

17. VOLKMANN, A. Ueber den Einfluss der Uebung auf das Erkennen raumlicher Distanzen. *Ber. d. Sachs. Ges. d. Wiss., math. phys. Abth.,* 1858, **10**, 38–69.

18. WHIPPLE, G. M. *Manual of mental and physical tests.* Part I. *Simpler processes.* Baltimore: Warwick and York, 1924.

(Received January 8, 1954)

Psychological Review
Vol. 62, No. 1, 1955

THE PRINCIPLE OF CONGRUITY IN THE PREDICTION OF ATTITUDE CHANGE

CHARLES E. OSGOOD AND PERCY H. TANNENBAUM
Institute of Communications Research, University of Illinois

The theoretical model presented in this paper, while not pretending to take account of all variables relating to attitude change, does attempt to cover those variables believed to be most significant with respect to the direction of change to be expected in any given situation. These variables are (*a*) existing attitude toward the source of a message, (*b*) existing attitude toward the concept evaluated by the source, and (*c*) the nature of the evaluating assertion which relates source and concept in the message. Predictions generated by the theory about the directions and relative amounts of attitude change apply to both sources and the concepts they evaluate.

UNDERLYING NOTIONS

Our work on attitude theory and measurement is an outgrowth of continuing research on experimental semantics, particularly the development of objective methods for measuring meaning (**4**, **5**). From this viewpoint, the *meaning* of a concept is its location in a space defined by some number of factors or dimensions, and *attitude* toward a concept is its projection onto one of these dimensions defined as "evaluative." In the factor analytic work we have done so far, the first and most heavily loaded factor is always one clearly identifiable as evaluative by the labels of the scales it represents, e.g., good-bad, fair-unfair, valuable-worthless, pleasant-unpleasant, and the like. This conception of attitude as a dimension or factor in total meaning has a number of implications, including those explored in the present paper. It implies, for example, that people having the same attitude toward a concept, such as NEGRO, may be sharply differentiated in terms of other dimensions of the semantic space (e.g., some perceiving NEGRO as powerful and active, others as weak and passive).[1]

Attitudes toward the various objects of judgment associated in messages must be measured in the same units if comparative statements about attitude change are to be made. There have been attempts to devise *generalized attitude scales* in the history of this field (cf. **6**, **7**), but if one is to judge by the criterion of acceptance and use, they have not been outstandingly successful. In applying the *semantic differential* (a label that has come to be applied to our measuring instrument), various objects of judgment, sources and concepts, are rated against a standard set of descriptive scales. To the extent that location on the evaluative dimension of the semantic differential is a reliable and valid index of attitude (as determined by correlation with other criteria), it is then necessarily a generalized attitude scale. We have some evidence for validity[2] and more is being obtained; reliability of the differential, particularly the evaluative dimension, is reasonably high, running in the .80's and .90's in available data.

---

[1] A study in progress exhibits precisely this phenomenon with respect to the concept NEGRO. Similar findings are evident with respect to THE CHURCH and CAPITAL PUNISHMENT.

[2] For example, the correlations between scores on the evaluative scales of the semantic differential and scores on the Thurstone scales on attitude toward THE CHURCH, NEGRO, and CAPITAL PUNISHMENT are .74, .82, and .81, respectively.

Another underlying notion about human thinking we have been exploring is that *judgmental frames of reference tend toward maximal simplicity.* Since extreme, "all-or-nothing" judgments are simpler than finely discriminated judgments of degree, this implies a continuing pressure toward polarization along the evaluative dimension (i.e., movement of concepts toward either entirely good or entirely bad allocations). We have evidence that extreme judgments have shorter latencies than more discriminative judgments (**5**), and that extreme judgments are characteristic of less intelligent, less mature, less well educated, or more emotionally oriented individuals (**8**). Furthermore, since assumption of identity is a simpler process than maintenance of distinction, this also implies a continuing pressure toward elimination of differences among concepts which are localized in the same direction of the evaluative framework. We have evidence that in the judging of emotionally polarized concepts all scales of judgment tend to rotate toward the evaluative, e.g., their correlations with good-bad tend to increase and therefore the relative loading on the evaluative factor tends to increase (**5**).

The most "simple-minded" evaluative frame of reference is therefore one in which a tight cluster of highly polarized and undifferentiated good concepts is diametrically opposed in meaning to an equally tight and polarized cluster of undifferentiated bad concepts. The same underlying pressure toward simplicity operates on any new or neutral concept to shift it one way or the other. For example, there is the tendency in American thinking, about which Pandit Nehru complains, requiring that India be either "for us or agin' us." This is, of course, the condition referred to by the general semanticists (e.g., Johnson, **2**) as a "two-valued orientation," and it is unfortunately characteristic of lay thinking in any period of conflict and emotional stress. The more sophisticated thinker, according to this view, should show less tendency to polarize, more differentiation among concepts, and thus greater relative use of factors other than the evaluative.

THE PRINCIPLE OF CONGRUITY

The principle of congruity in human thinking can be stated quite succinctly: *changes in evaluation are always in the direction of increased congruity with the existing frame of reference.* To make any use of this principle in specific situations, however, it is necessary to elaborate along the following lines: When does the issue of congruity arise? What directions of attitude change are congruent? How much stress is generated by incongruity and how is it distributed among the objects of judgment?

The issue of congruity. Each individual has potential attitudes toward a near infinity of objects. It is possible to have varying attitudes toward diverse concepts without any felt incongruity or any pressure toward attitude change, as long as no association among these objects of judgment is made. As anthropologists well know, members of a culture may entertain logically incompatible attitudes toward objects in their culture (e.g., ancestor worship and fear of the dead) without any stress, as long as the incompatibles are not brought into association. The issue of congruity arises whenever a message is received which relates two or more objects of judgment via an assertion.

The simplest assertion is merely a *descriptive statement:* "Chinese cooking is good," "Jefferson was right," "This neurotic modern art." To the extent that the evaluative location of a particular qualifier differs from that of the thing qualified, there is generated some pressure toward congruity. Similar pressure is generated by ordinary *statements of classification:* "Senator McCarthy is a Catholic," "Tom is an ex-con," "Cigarettes contain nicotine." To the extent that the evaluative locations of instance and class are different, some pressure toward congruity exists. A more complex situation is that in which *a source makes an assertion about a concept:* "University President Bans Research on Krebiozen"; "Communists like strong labor unions." This is the most commonly studied situation, and one for which we have some empirical data against which to test our hypotheses. Assertions may be explicit linguistic statements of evaluation or implicit behavioral, situational statements. A newsphoto of Mrs. Roosevelt smiling and shaking hands with a little colored boy is just as effective in setting up pressures toward congruity as a verbal statement on her part.

Directions of congruence and incongruence. To predict the direction of attitude change from this general principle it is necessary to take into account simultaneously the existing attitudes toward each of the objects of judgment prior to reception of the message and the nature of the assertion which is embodied in the message. Attitudes can be specified as favorable (+), neutral (0), and unfavorable (−). Assertions can be specified as positive or associative (+) or negative or disassociative (−). They may also, of course, include evaluative loading (e.g., when X denounces Y, we have both a disassociative assertion and negative evaluation of Y). When attitudes toward both objects of judgment are polar, the nature of the assertion determines congruence or incongruence. For EISENHOWER (+) to come out in favor of FREEDOM OF THE PRESS (+) is, of course, congruent with the existing frame of reference of most people in this country, but for THE DAILY WORKER (−) to speak in favor of FREEDOM OF THE PRESS (+) is attitudinally incongruent. In this simplest of states in which human thinking operates, sources we like should always sponsor ideas we like and denounce ideas we are against, and vice versa.

When the existing attitude toward one of the objects of judgment is neutral and the other polar, we must speak of what directions *would be* congruent. If, for example, a favorable source like EISENHOWER were to make a favorable assertion about the MINISTER FROM SIAM (a neutral notion to most of us), it would be congruent *if* the latter were also favorable—hence pressure is generated toward attitude change in this direction. If PRAVDA (−) sponsors GRADUAL DISARMAMENT (0), the pressure is such as to make the relatively neutral notion of disarmament less favorable; similarly, if a PROFESSOR (0) as a source favors PREMARITAL SEXUAL RELATIONS (−) as making for better marriages, it is the PROFESSOR that becomes less favorable (this is not unlike the "guilt by association" technique). Conversely, for our neutral PROFESSOR (0) to speak out against MORAL DEPRAVITY (−) must have the effect of raising his esteem (this is the familiar "I am against sin" technique).

When both objects of judgment are neutral, there is no question of congruity between them, and movement is determined solely by the nature of the assertion, i.e., this becomes a case of simple qualification or classification. If MR. JONES denounces MR. SMITH, neither of whom is known, there is presumably some negative pressure on MR. SMITH by virtue of the sheer devaluation of "being denounced." Since the evaluation applies to the concept and not the source, the effect should be chiefly upon the concept. We shall find evidence for such an "assertion effect" in the available data.

We may now make a general statement governing the direction of congruence which will hold for any object of judgment, source or concept, and any type of assertion.

1. *Whenever one object of judgment is associated with another by an assertion, its congruent position along the evaluative dimension is always equal in degree of polarization* (d) *to the other object of judgment and in either the same (positive assertion) or opposite (negative assertion) evaluative direction.*

This is to say that we have attitude scores toward two objects of judgment, OJ_1 and OJ_2, and to each of these scores we assign a value, d, which represents the degree of polarization of that attitude. Thus we have d_{OJ1} and d_{OJ2}. Since the measuring instrument which has been used in our quantitative work so far (the semantic differential) treats the evaluative dimension as a 7-step scale with "4" defined as the neutral point, we have three degrees of polarization in each direction, i.e., $+3$, $+2$, $+1$, 0, -1, -2, -3. Given OJ_1 and OJ_2 associated with one another through either a positive ($OJ_1A^+OJ_2$) or negative ($OJ_1A^-OJ_2$) assertion, we define the congruent position (C) of either object of judgment as follows:

If $OJ_1A^+OJ_2$, then

$$dc_{OJ_1} = d_{OJ_2}, \qquad [1]$$

$$dc_{OJ_2} = d_{OJ_1}. \qquad [2]$$

If $OJ_1A^-OJ_2$, then

$$dc_{OJ_1} = -\,d_{OJ_2}, \qquad [3]$$

$$dc_{OJ_2} = -\,d_{OJ_1}. \qquad [4]$$

Figure 1 provides some graphic illustrations. In example 1, we have a positive assertion (indicated by the + on the bar connecting source and concept) associating two equally favorable objects of judgment; in this situation maximum congruity already exists. In all the other illustrations given, the existing positions are not those of maximum congruence, and those positions which would be maximally congruent for each object of judgment are shown by dashed circles. In situation 3, for example, a congruent source would be at -2 and a congruent concept would be at $+3$, given the favorable assertion between two items of opposite sign.

Magnitude and distribution of pressure toward congruity. Knowing the existing locations of maximum congruence under the given conditions (by applying Principle 1), it becomes possible to state the amount and

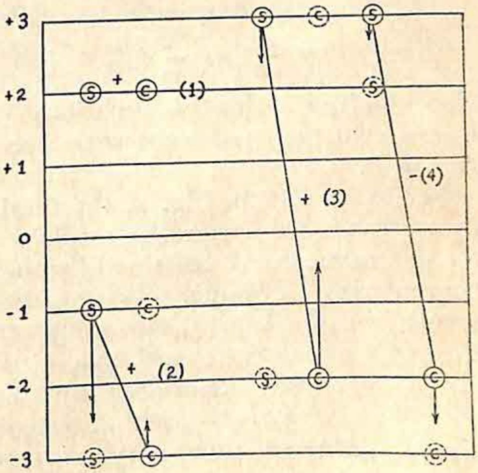


FIG. 1. Graphic examples of four situations in which a source (s) makes an assertion (+ or −) about a concept (c). Positions of maximum congruity are indicated by dashed circles; predicted changes in attitude are indicated by arrows. See discussion in text.

direction of application of total available pressure toward congruity.

2. *The total available pressure toward congruity* (P) *for a given object of judgment associated with another by an assertion is equal to the difference, in attitude scale units, between its existing location and its location of maximum congruence along the evaluative dimension; the sign of this pressure is positive* (+) *when the location of congruence is more favorable than the existing location and negative* (−) *when the location of congruence is less favorable than the existing location.*

That is,

$$P_{OJ_1} = dc_{OJ_1} - d_{OJ_1}, \qquad [5]$$

$$P_{OJ_2} = dc_{OJ_2} - d_{OJ_2}. \qquad [6]$$

Therefore, substituting from equations 1 through 4,

if $OJ_1A^+OJ_2$,

$$P_{OJ_1} = d_{OJ_2} - d_{OJ_1}, \qquad [7]$$

$$P_{OJ_2} = d_{OJ_1} - d_{OJ_2}. \qquad [8]$$

If $OJ_1A^-OJ_2$,

$$P_{OJ_1} = -\,d_{OJ_2} - d_{OJ_1}, \qquad [9]$$

$$P_{OJ_2} = -\,d_{OJ_1} - d_{OJ_2}. \qquad [10]$$

The resulting signs of equations 7 through 10 then represent the direction of P.

For example 2 in Fig. 1, the total pressure toward congruity available for the source is −2 units and for the concept is +2 units. As can be seen by inspection of these examples, the total pressures toward congruity for both objects associated by an assertion are always equal in magnitude, although they may be the same or different in sign (i.e., $|P_{OJ_1}| = |P_{OJ_2}|$). The upper figures in each cell of Table 2 give the total pressures and directions of application for all possible relations among sources and concepts and both types of assertions. These computations are based upon the assumption of a 7-step scale with three degrees of polarization possible in each evaluative direction; they may be treated as general index numbers.

The third principle with which we shall operate incorporates the empirical generalization that intense attitudes are more resistant to change than weakly held ones (cf., **1**, **3**, **9**), but does so in a way which generates more detailed predictions.

3. *In terms of producing attitude change, the total pressure toward congruity is distributed between the objects of judgment associated by an assertion in inverse proportion to their separate degrees of polarization.*

In other words, relatively less polarized objects of judgment, when associated with relatively more polarized objects of judgment, absorb proportionately greater amounts of the pressure toward congruity, and consequently change more.

Applying Principle 3 above, it is possible to predict relative attitude change according to the following formulas:

$$AC_{OJ_1} = \frac{|d_{OJ_2}|}{|d_{OJ_1}| + |d_{OJ_2}|}\, P_{OJ_1}, \qquad [11]$$

$$AC_{OJ_2} = \frac{|d_{OJ_1}|}{|d_{OJ_1}| + |d_{OJ_2}|}\, P_{OJ2}, \qquad [12]$$

where AC refers to attitude change; where d_{OJ_1} and d_{OJ_2} are taken at their absolute values regardless of sign, and where P_{OJ_1} and P_{OJ_2} are determined from equations 7 through 10. Thus, the sign of the right-hand side of the equation is always that of the particular P_{OJ} under consideration, and thus represents the direction of change. In example 1 in Fig. 1, there is no pressure and hence no change. In the other examples solid arrows indi-

TABLE 1

Predicted Attitude Change for Concept as a Function of Original Locations of Both Source and Concept—Positive Assertion

(Uncorrected for incredulity)

Original Attitude Toward Source	Original Attitude Toward Concept						
	+3	+2	+1	0	−1	−2	−3
+3	0.0	+0.6	+1.5	+3.0	+3.0	+3.0	+3.0
+2	−0.4	0.0	+0.7	+2.0	+2.0	+2.0	+2.0
+1	−0.5	−0.3	0.0	+1.0	+1.0	+1.0	+1.0
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
−1	−1.0	−1.0	−1.0	−1.0	0.0	+0.3	+0.5
−2	−2.0	−2.0	−2.0	−2.0	−0.7	0.0	+0.4
−3	−3.0	−3.0	−3.0	−3.0	−1.5	−0.6	0.0

cate the direction and magnitude of predicted change. In example 2, the source must absorb twice as much pressure as the more polarized concept, and in a negative rather than a positive direction (e.g., BULGARIA −1 sponsors HATE CAMPAIGN −3). The unexpected prediction here, that the even more unfavorable concept, HATE CAMPAIGN, actually becomes a little less unfavorable under these conditions, derives directly from the theoretical model and will be discussed later.

If we associate OJ_1 with the source of an assertion, and OJ_2 with the concept, the numbers in Table 1 represent the results of applying formula 12 to the prediction of attitude change toward the concept for all combinations of original attitude toward source and concept when the assertion is positive.

Note that for all *incongruous relations* the predicted change is constant for a given original attitude toward source (upper right and lower left corners of matrix)—a highly favorable source favoring a negative concept produces just as much attitude change when that concept is −3 as when it is −1. This prediction assumes complete credulity of the message on the part of the receiver, a condition that exists only rarely, in all probability, for incongruous messages. Certainly, when presented with the incongruous message, EISENHOWER sponsors COMMUNISM, in an experimental situation, very few subjects are going to give it full credence. If we are going to make predictions, it is apparent that the variable of credulity must be taken into account.

4. *The amount of incredulity produced when one object of judgment is associated with another by an assertion is a positively accelerated function of the amount of incongruity which exists and operates to decrease attitude change, completely eliminating change when maximal.*

Since incongruity exists only when similarly evaluated concepts are associated by negative assertions or when oppositely evaluated concepts are

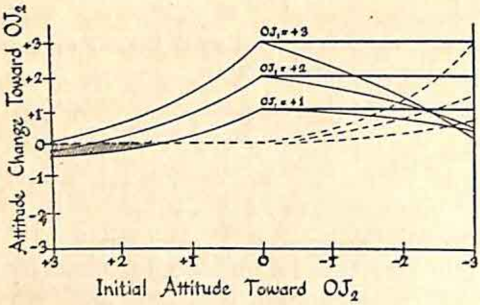


FIG. 2. Predicted change in attitude toward OJ_2 and correction for incredulity (postulated incredulity indicated by dashed lines).

associated by positive assertions, the correction for incredulity is limited to the upper right and lower left corners of the matrix in Table 1. Within these situations, the *amount* of this correction is assumed to increase with the degree of incongruity, e.g., with the total pressure toward congruity available. It is assumed that no incongruity, and hence no incredulity, can exist where one of the objects of judgment is neutral, e.g., EISENHOWER may come out either for or against a neutral concept like ST. LAWRENCE WATERWAY without the issue of incredulity arising. It is realized, of course, that other factors than those discussed here may affect incredulity.

Figure 2 provides graphic illustration of the corrections made for incredulity. The original curves, level for the neutral point and beyond, derive from the upper right corner of Table 1 and represent three degrees of favorable original attitude toward the source (OJ_1). The dashed lines represent postulated incredulity, positively accelerated functions of total pressure toward congruity. The shape of this function is, of course, based on pure hunch and will probably have to be modified; it simply seems reasonable that a person's tendency to reject a message will be relatively much less for slightly incongruous relations (e.g., EISENHOWER +3 praises BULGARIANS −1) than for grossly incongruous ones (e.g., EISENHOWER +3 praises COMMUNISM −3). The light solid curves represent the result of subtracting incredulity functions from predicted attitude changes. The corrected values are given in Table 2. The same corrections, with appropriate regard to sign, apply to the lower left corner of Table 1 and are made in Table 2. For incongruous situations, then, the formulas for predicting attitude change become:

$$AC_s = \frac{|d_c|}{|d_s| + |d_c|} P_s \mp i, \quad [13]$$

$$AC_c = \frac{|d_s|}{|d_s| + |d_c|} P_c \mp i, \quad [14]$$

where the sign of the second factor (i = correction for incredulity) is always opposite to that of the first factor, i.e., of P, and thus serves to diminish the effect. The notations s and c refer, of course, to source and concept, respectively.

It is to be emphasized that formulas 13 and 14 apply only to situations of incredulity, or, to put it another way, $i = 0$ for all credulous situations. The precise nature of the i function remains to be empirically determined. The curve we have assumed for i approximates the function $i = a(d_s^2 + b)(d_c^2 + d_c)$, where the constants a and b are 1/40 and 1 respectively. It is clear, however, that $i = f(d_s, d_c)$, and hence the attitude change is still a function of the two degrees of polarization and of direction of the assertion.

Incredulity, to the extent it is present, will not only operate to "damp" changes in attitude but should also appear in expressions of disbelief and rationalization. This makes it possible to ascertain the incredulity function independent of attitude change. A proposed experiment along these lines might be as follows: If subjects are presented with a number of messages and told that some are valid and others faked, we would expect the frequency of "fake" judgments to be some increasing function of the measured amount of incongruity (e.g., in terms of locations of original attitudes toward the associated objects of judgment). It is the shape of this function in which we would be particularly interested. If the same subjects were then assured that certain of the "fake" messages were actually valid, we would expect to record rationalizations and other attempts to interpret the message without modifying the evaluative frame of reference; e.g., told that RUSSIA is actually sponsoring a PEACE CONFERENCE, the subject is likely to rationalize this event as some subterfuge on the part of the Soviets in the Cold War. Use of the same messages on another group of subjects in a standard pre- and postmessage attitude change design should serve to test the prediction that attitude change for incongruous assertions is damped in proportion to the degree of incredulity produced.

This independence of credulity as a variable also means that it should be possible to approximate the attitude change values in Table 1 (uncorrected) under special conditions where credulity is made more probable. For example, if EISENHOWER

TABLE 2

TOTAL PRESSURE TOWARD CONGRUITY (UPPER NUMBERS) AND PREDICTED CHANGES IN ATTITUDE (LOWER NUMBERS) AS CORRECTED FOR INCREDULITY*

(Positive assertion)**

Initial Location of OJ_2	Initial Location of OJ_1							
	+3	+2	+1	0	−1	−2	−3	*M*
+3	0 0.0	+1 +0.6	+2 +1.5	+3 +3.0	+4 +2.6	+5 +1.5	+6 0.0	+1.3
+2	−1 *−0.4*	0 0.0	+1 +0.7	+2 +2.0	+3 +1.8	+4 +1.3	+5 +0.5	+0.8
+1	−2 *−0.5*	−1 *−0.3*	0 0.0	+1 +1.0	+2 +0.9	+3 +0.7	+4 +0.4	+0.3
0	−3 0.0	−2 0.0	−1 0.0	0 0.0	+1 0.0	+2 0.0	+3 0.0	0.0
−1	−4 −0.4	−3 −0.7	−2 −0.9	−1 −1.0	0 0.0	+1 *+0.3*	+2 *+0.5*	−0.3
−2	−5 −0.5	−4 −1.3	−3 −1.8	−2 −2.0	−1 −0.7	0 0.0	+1 *+0.4*	−0.8
−3	−6 0.0	−5 −1.5	−4 −2.6	−3 −3.0	−2 −1.5	−1 −0.6	0 0.0	−1.3
/*M*/	0.3	0.6	1.1	1.7	1.1	0.6	0.3	

---

* The *OJ* whose change is being predicted, either source or concept, is always OJ_1 and the other is OJ_2.

** When dealing with negative assertions, reverse the sign of OJ_2 and look in that row.

were actually to invite important COMMUNIST PARTY OFFICERS to a friendly dinner at the White House, the effects on devaluating EISENHOWER and vice versa should be extreme. The greater effectiveness of "event" as compared to "word" propaganda follows directly.

The lower numbers in each cell in Table 2 represent the predicted magnitudes and directions of attitude change for all combinations of original attitude—for SOURCES or CONCEPTS and for either positive assertions or negative assertions (see instructions in table footnotes)—as computed by applying formulas 11 and 12, or 13 and 14 in the case of incongruous assertions. Let us take illustration 2 in Fig. 1 as an example: the original attitudes in this case are -1 (source) and -3 (concept), so we will be concerned with the cell defined by these values. Looking first for SOURCE CHANGE (e.g., SOURCE as OJ_1), we find a total pressure toward congruity of -2 (upper figure in cell) and a predicted attitude change of -1.5; looking then for CONCEPT CHANGE (e.g., CONCEPT as OJ_1), we find a total pressure of $+2$ and a predicted attitude change of $+0.5$. If this were the message BULGARIA praises COMMUNISM, we might expect a considerable increase in unfavorableness toward BULGARIA and a slight decrease in unfavorableness toward COMMUNISM. The reader may check the other illustrations against Table 2 if he wishes.

Predictions about attitude change are assumed to hold for any situation in which one object of judgment is associated with another by an assertion. In the special source-concept situation, we must take one additional factor into account—the fact that the *assertion itself*, whether positive or negative, typically applies to the concept rather than to the source. When X praises Y, the favorable effect of "praise" applies chiefly to Y; when X denounces Y, similarly, the unfavorable effect of "denounce" applies chiefly to Y. In other words, we must add to the equation predicting attitude change for concepts a constant ($\pm A$) whose sign is always the same as that of the assertion. In the data we have available, the existence of such an assertion constant applying to the concept but not the source is clearly evident.

A TEST OF THE CONGRUITY PRINCIPLE[3]

On the basis of a pretest of 36 potential objects of judgment, three source-concept pairs were selected which met the criteria of (*a*) approximately equal numbers of subjects holding favorable, neutral, and unfavorable original attitudes toward them, and (*b*) lack of correlation between attitude toward the source and the concept making up each pair. The three source-concept pairs finally selected were: LABOR LEADERS with LEGALIZED GAMBLING, CHICAGO TRIBUNE with ABSTRACT ART, and SENATOR ROBERT TAFT with ACCELERATED COLLEGE PROGRAMS. Another group of 405 college students was given a *before-test*, in which the 6 experimental concepts along with 4 "filler" concepts were judged against a form of the semantic differential including 6 scales highly loaded on the evaluative factor. The sum of ratings on these six 7-step scales constituted the attitude score for each concept, these scores ranging from 6 (most unfavorable) to 42 (most favorable). Five weeks later the same subjects were given highly realistic newspaper stories including positive or negative assertions involving the experimental source-concept pairs.

---

[3] This experiment is described in detail in a separate report by one of the authors (9).

TABLE 3

PREDICTED (UPPER VALUES IN CELLS) AND OBTAINED (LOWER VALUES IN CELLS) CHANGES IN ATTITUDE

Original Attitude Toward Source	Positive Assertions			Negative Assertions		
	Original Attitude Toward Concept			Original Attitude Toward Concept		
	+	0	−	+	0	−
			Source Changes			
+	+0.2 + 25	0.0 + 16	−1.1 − 42	−1.1 − 45	0.0 + 1	+0.2 + 34
0	+2.0 +150	0.0 + 25	−2.0 − 94	−2.0 − 68	0.0 + 17	+2.0 + 96
−	+1.1 + 49	0.0 + 13	−0.2 − 7	−0.2 − 33	0.0 − 3	+1.1 + 34
			Concept Changes			
+	+0.2 +51	+2.0 +245	+1.1 +107	−1.1 − 88	−2.0 −180	−0.2 − 39
0	0.0 + 39	0.0 + 80	0.0 + 48	0.0 − 72	0.0 − 79	0.0 − 34
−	−1.1 − 24	−2.0 − 52	−0.2 − 10	+0.2 + 19	+2.0 + 22	+1.1 + 16

Immediately afterward the subjects were given the *after-test*, again judging the same concepts against the semantic differential.

Original attitudes toward each source and concept were determined from the before-test scores, subjects being distributed into nine cells, s_+c_+, s_+c_0, s_+c_-, s_0c_+, etc. Attitude change amounts, for both source and concept, were obtained by subtracting the before-test score from the after-test score for each subject, a positive value thereby indicating increased favorableness.

Table 3 compares predicted attitude change scores (upper number in each cell) with obtained attitude change scores (lower number in each cell) for both sources and concepts and for both positive and negative assertions. The predicted values represent the algebraic mean of the attitude change scores in appropriate cells of Table 2 (e.g., the value for s_+c_+ with a positive assertion equals the average for the nine cells in the upper left corner); the obtained values represent the total attitude change scores, summed algebraically, for 45 subjects (15 subjects on each of 3 stories) on 6 evaluative scales. The reason for the gross difference in absolute magnitudes of predicted and obtained scores is therefore that the former are expressed in scale units and the latter in group totals. The general correspondence between predicted and obtained directions of attitude change is apparent from inspection of Table 3. In every case predicted positive changes (+) and predicted negative changes (−) show corresponding signs in the obtained data, and predicted lack of change (0) generally yields obtained changes of small magnitude.

The predictions obviously hold better for source changes than for concept changes, and it will be recalled that it was also predicted that *an assertion constant* ($\pm A$) would apply to the concept but not the source. This would mean that for comparable situations (e.g., s_+c_0 vs. s_0c_+, s_0c_- vs. s_-c_0, etc.) concept changes should be more in the favorable direction than source changes for positive assertions and more in the negative direction for negative assertions. Table 4 provides a test of this prediction. As can be seen, when comparable conditions for source and concept changes are arranged, the differences in magnitudes of attitude change are regularly positive for positive assertions (concept changes more toward favorable direction) and regularly negative for negative assertions (concept changes more toward unfavorable direction). With 17 of the 18 values in the predicted direction, this is obviously significant.

A rough estimate of the size of this constant can be obtained from the average difference between source and concept changes for comparable situations: it turns out to be $A = \pm 46$ in total change score or $A = .17$ in units of the 7-step attitude scale employed. In other words, under the general conditions of this experiment, the assertion constant applied to concept changes equals about 1/6 of a scale unit.

That the magnitudes of attitude changes as well as their directions tend to follow predictions is also evident by inspection of Table 3. This can be seen more clearly in the correlation plot between predicted changes and obtained changes given as Fig. 3. The predicted values are here treated as categorical and the obtained as continuous, and the latter have been corrected for the assertion constant by adding 46 to concept changes with negative assertions and subtracting 46 from concept changes with positive assertions. The correlation between predicted and obtained changes in attitude is high ($r = .91$).

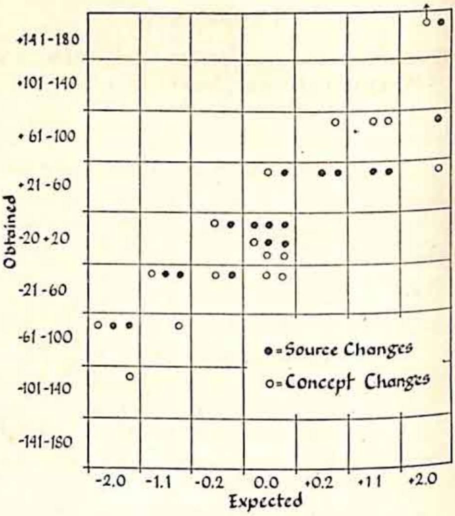


FIG. 3. Correlation plot between predicted and obtained attitude changes after corrections for incredulity and for constant effect of assertion.

TABLE 4

EFFECTS OF ASSERTION ITSELF (A)

Source	Concept	Difference (A)
Positive Assertions (Predicted that A is +)		
s_+c_+ + 25	s_+c_+ + 51	+ 26
s_0c_+ +150	s_+c_0 +245	+ 95
s_-c_+ + 49	s_+c_- +107	+ 58
s_+c_0 + 16	s_0c_+ + 39	+ 23
s_0c_0 + 25	s_0c_0 + 80	+ 55
s_-c_0 + 13	s_0c_- + 48	+ 35
s_+c_- − 42	s_-c_+ − 24	+ 18
s_0c_- − 94	s_-c_0 − 52	+ 42
s_-c_- − 7	s_-c_- − 10	− 3
Negative Assertions (Predicted that A is −)		
s_+c_+ − 45	s_+c_+ − 88	− 43
s_0c_+ − 68	s_+c_0 −180	−112
s_-c_+ − 33	s_+c_- − 39	− 6
s_+c_0 + 1	s_0c_+ − 72	− 73
s_0c_0 + 17	s_0c_0 − 79	− 96
s_-c_0 − 3	s_0c_- − 34	− 31
s_+c_- + 34	s_-c_+ + 19	− 15
s_0c_- + 96	s_-c_0 + 22	− 74
s_-c_- + 34	s_-c_- + 16	− 18

A number of corollaries derive from the congruity principle, some of which can be tested against Tannenbaum's data and others of which cannot. They are as follows:

1. *Shifts in evaluation always tend toward equalization of the degrees of polarization of the objects of judgment associated by an assertion.* If two unequally polarized concepts are associated, the less polarized one becomes more so and the more polarized less so; if a neutral concept is associated with a polarized one, it always becomes more polarized. In Tannenbaum's data this means that the less polarized object of judgment (neutral) should always change more than the more polarized object of judgment (plus or minus); this holds for all

TABLE 5

COMPARISON OF ATTITUDE CHANGES FOR NEUTRAL AND POLAR OBJECTS OF JUDGMENT

Situation	Change for Neutral	Change for Polar
Positive Assertions		
s_0c_+	+150 (*s*)	−7 (*c*)
s_+c_0	+199 (*c*)	+16 (*s*)
s_-c_0	−98 (*c*)	+13 (*s*)
s_0c_-	−94 (*s*)	+2 (*c*)
Negative Assertions		
s_0c_+	−68 (*s*)	−26 (*c*)
s_+c_0	−134 (*c*)	+1 (*s*)
s_-c_0	+48 (*c*)	−3 (*s*)
s_0c_-	+96 (*s*)	+12 (*c*)

relevant conditions except one (s_0c_+, negative assertion), and even here when correction for the assertion constant is made. The comparisons shown in Table 5, where concept changes have been corrected for the assertion constant, clearly substantiate this prediction—in every case the neutral member shows a large shift and the polar member a small shift.

2. *When the total pressure toward congruity is constant, it is easier to make an object of judgment more polarized than less so.* That this prediction follows from the theory can be seen in Table 2 by comparing magnitudes of attitude change within the same row or column (for a given location of source or concept) when the total pressure is the same in amount but opposite in sign—amount of attitude change is always larger in the same direction as the sign of the row or column. This cannot be checked in Tannenbaum's data because he does not differentiate between degrees of original attitude in the same direction. It would be expected, however, from our general notion that evaluative frames of reference tend toward maximum simplicity.

3. *Attitude change toward an object of judgment is an inverse function of intensity of original attitude toward that object.* That weakly held attitudes are more susceptible to change is a widely held notion, but it is only valid, according to the present theory, for the *average* of all degrees of attitude toward the other object with which a given one is associated *regardless of sign* (the absolute means, $|M|$, given at the bottom of Table 2) or for maximally polarized attitudes toward the other object (+3 and −3 rows in Table 2). The "law" definitely does not hold for other degrees of attitude toward the second object of judgment, as can be seen by inspection of this table. The data in Table 6 compare predicted values (absolute means at bottom of Table 2) with those obtained by Tannenbaum, both being expressed in attitude scale units in this case. The close correspondence in trend is apparent, and the obtained trend is statistically significant. The difference in absolute magnitude presumably represents the limited effect of a single message upon attitude change. The theoretical model as developed so far takes account of neither learning via suc-

TABLE 6

ATTITUDE CHANGE TOWARD AN OBJECT OF JUDGMENT AS A FUNCTION OF ORIGINAL ATTITUDE TOWARD THAT OBJECT ITSELF

	Original Attitude Toward Object of Judgment		
	+	0	−
Predicted	0.6	1.7	0.6
Obtained (Source)	0.2	0.5	0.2
Obtained (Concept)	0.3	0.7	0.2

TABLE 7

ATTITUDE CHANGE TOWARD ONE OBJECT OF JUDGMENT AS A FUNCTION OF ORIGINAL ATTITUDE TOWARD THE OTHER OBJECT OF JUDGMENT

Assertions	Original Attitude Toward Other Object		
	+	0	−
Positive			
Predicted	+0.8	0.0	−0.8
Obtained (Source)	+0.3	0.0	−0.2
Obtained (Concept)	+0.3	0.0	−0.3
Negative			
Predicted	−0.8	0.0	+0.8
Obtained (Source)	−0.2	0.0	+0.2
Obtained (Concept)	−0.2	−0.1	+0.2

cessive messages nor of intensity of assertions.

4. *Attitude change for a given object of judgment in the direction of the assertion is an approximately linear function of the favorableness of the original attitude toward the other object of judgment with which it is associated.* The more favorable the attitude toward a source, the greater the effect of a positive assertion on raising attitude toward the concept and the greater the effect of a negative assertion upon lowering attitude toward the concept. Strongly unfavorable sources have just the opposite effects. The same statements hold for changes in attitude toward sources when original attitudes toward concepts are varied. Table 7 compares predicted and obtained values. Again, the generally lower levels of obtained changes as compared with predicted changes are presumably due to the limited effects of a single message. Changes in attitude have been corrected for the assertion constant, here as in Table 6. The obtained functions are in the direction predicted, and their trend is statistically significant.

5. *Whenever a congruent assertion associates two differently polarized objects of judgment, and neither of them is neutral, the more polarized object of judgment becomes less so.* This deduction includes the rather paradoxical situations noted earlier in this paper in which, for example, a highly favorable source comes out in favor of somewhat less favorable concept and becomes slightly less favorable itself in doing so, according to theory. The locations in Table 2 where this prediction arises are italicized. Here we are forced to predict, for example: when EISENHOWER +3 praises GOLFING +1, he loses a little prestige while giving a big boost to the concept; when EISENHOWER +3 denounces COMIC BOOKS −1, he may make the latter considerably more unfavorable, but he loses a little ground himself in the exchange. It is as if a highly favorable source should only favor equally good things or be against extremely bad things and a highly unsavory concept should be only sponsored by equally unsavory sources or condemned by highly noble sources.

If such an effect could be demonstrated, it would be convincing evidence for the whole theory. Tannenbaum's study provides only a partial test for this phenomenon in that the experimental situation probably lacked the necessary sensitivity to get at the very minimal changes predicted. However, of 38 cases that met the necessary conditions, 21 (55.3 per cent) showed relatively small changes in the predicted direction, 15 (39.5 per cent) showed no change, and only 2 (5.3 per cent) changed in the opposite direction.

SUMMARY

This paper describes a general theory of attitude change which takes into account original attitude toward the source of a message, original attitude toward the concept evaluated by the source, and the nature of the evaluative assertion. Predicted changes in attitude toward both source and concept are based upon the combined operation of a principle of congruity, a principle of susceptibility as a function of polarization, and a principle of resistance due to incredulity for incongruous messages. Comparison of predictions with data obtained in a recent experiment provides a test of the theory. No attempt has been made to integrate this particular theoretical model with more general psychological theory, and we feel no urge at this time to attempt such detailed translations. We are, of course, aware that there are many variables other than those considered here which contribute to attitude change.

REFERENCES

1. BIRCH, H. G. The effect of socially disapproved labeling upon well-structured attitudes. *J. abnorm. soc. Psychol.*, 1945, **40**, 301–310.
2. JOHNSON, W. The communication process and general semantic principles. In W. Schramm (Ed.), *Mass communications.* Urbana: Univer. of Illinois Press, 1949. Pp. 261–274.
3. KLAPPER, J. T. The effects of mass media. Mimeographed manuscript, Bureau of Applied Social Research, Columbia Univer., 1949.
4. OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 197–237.
5. OSGOOD, C. E. Report on development and application of the semantic differential. Unpublished manuscript, Institute of Communications Research, Univer. of Illinois, 1953. Privately distributed.
6. REMMERS, H. H. (Ed.) Generalized attitude scales: studies in socio-psychological measurements. *Bull. Purdue Univer. Stud. higher Educ.*, 1934, **36**, No. 4.
7. REMMERS, H. H., & SILANCE, E. B. Generalized attitude scales. *J. soc. Psychol.*, 1934, **5**, 298–312.
8. STAGNER, R., & OSGOOD, C. E. An experimental analysis of a nationalistic frame of reference. *J. soc. Psychol.*, 1941, **14**, 389–401.
9. TANNENBAUM, P. H. Attitudes toward source and concept as factors in attitude change through communications. Unpublished doctor's dissertation, Univer. of Illinois, 1953.

(Received January 15, 1954)

Psychological Review
Vol. 62, No. 1, 1955

A MODEL OF SHORT- AND LONG-RUN MECHANISMS INVOLVED IN PRESSURES TOWARD UNIFORMITY IN GROUPS

HERBERT A. SIMON AND HAROLD GUETZKOW[1]
Graduate School of Industrial Administration, Carnegie Institute of Technology

Leon Festinger and his associates (2) have stated and tested a number of propositions about communication processes in small groups. This paper carries forward the synthesis of these propositions into an interrelated system, differentiating short-run mechanisms from those involved in the long run. First, we shall formalize some of Festinger's propositions and organize them into a model. Then, we shall examine two of the empirical studies in their relation to the mechanisms involved in pressures toward uniformity in groups.

1. The Postulated Aggregative Model

Some of Festinger's hypotheses deal with aggregative relations—that is, they are concerned with the members of a group taken as a whole. Others focus on the behaviors of particular individuals in the group.[2] Let us confine our work for the present to his aggregative propositions.

The Aggregative Variables. Five of Festinger's hypotheses—those numbered 1a, 1b, 1c, 3a, and 3b—are stated in terms of six aggregative variables, in addition to time.

$D(t)$: The perceived *discrepancy* of opinion on an issue among members of a group at time t;

$P(t)$: *Pressure* upon members of the group to communicate with each other at time t;

$L(t)$: Receptivity (*listening*) of members of the group to influence by communications from other members at time t;

$C(t)$: Strength of the attraction of individuals to the group (*cohesiveness*) at time t;

$U(t)$: Pressure felt by the group to achieve *uniformity* of opinion, i.e., to reduce perceived discrepancy of opinion at time t;

R: *Relevance* of the issue to the group. This variable appears as a parameter, and hence as constant in time.

In constructing the model we shall assume that each variable is some kind of average or aggregate for members of the group. For example, D might be measured by locating the opinions of group members on a scale, attaching numbers to scale positions and calculating the standard deviation of the members' opinions in terms of these numbers. Even the intervening variables, although not directly measured, can be thought of as averages of the values for individual members. We will assume that the magnitude of each variable at time t for a given group can be represented by a real variable, using the latter term in its mathematical sense.

Representation of the Hypotheses by Equations. Festinger's aggregative propositions (**2**, pp. 273–277) state relationships between variables, as follows:

---

[1] An earlier version of this paper was presented to the Summer Seminar on Design of Experiments on Decision Processes at Santa Monica in July, 1952. In revising the paper we have had the benefit of valuable comments from Leon Festinger, James Coleman, and Chris Argyris.

[2] A model handling the deviate member has been developed and will be published by the *British Journal of Psychology* under the title "Some Mechanisms Involved in Group Pressures Upon Deviate-Members."

Number of the Proposition	Variables Interrelated
1a	D and P
1b	R and P
1c	C and P
3a	Change(s) in D and U
3b	Change(s) in D and C

We could now translate each of these five propositions into a corresponding mathematical relationship. This would not do justice, however, to their meaning, and might fail to represent adequately the dynamic mechanisms that Festinger implicitly postulates. Instead, we shall set forth a model of five equations which, we believe, constitutes a reasonable interpretation of the mechanisms. Then we shall show the relation of Festinger's explicit propositions to the system we have constructed. The five equations comprising the model are as follows:

$$\frac{dD}{dt} = f[P(t), L(t), D(t)],^{3} \qquad [1.1]$$

$$P(t) = P[D(t), U(t)], \qquad [1.2]$$

$$L(t) = L[U(t)], \qquad [1.3]$$

$$\frac{dC}{dt} = g[D(t), U(t), C(t)], \qquad [1.4]$$

$$U(t) = U[C(t), R]. \qquad [1.5]$$

Note that two of the equations, 1.1 and 1.4, postulate a process of adjustment, of D and C, respectively, that takes place gradually over time. The remaining equations, 1.2, 1.3, and 1.5, make P, L, and U, respectively, change instantaneously with the variables on which they depend. The term "instantaneously" need not, of course, be interpreted literally, but may be taken to mean that the mechanisms represented by these three equations bring about more rapid adjustment in the dependent variable than the mechanisms in the other two. This part of the system enables us to make clear distinctions between the short-run and long-run mechanisms involved in pressures toward uniformity. While the equations permit the values of the variable to change through time, the forms of the equations themselves are assumed to be independent of the length of time the group is in existence.

The propositions are stated so as to involve only *ordinal* and not *cardinal* properties of the variables. This is important, since in the present stage of development of operational definition of the variables the scale units are arbitrary. We can observe that group A is more cohesive than group B, but not that A is twice as cohesive as B, or that the cohesiveness of A exceeds the cohesiveness of B by a greater amount than the cohesiveness of C exceeds the cohesiveness of D. It should be stressed that our treatment does not impose any requirements of measurability not already present, implicitly or explicitly, in Festinger's verbal theory. The equations make statements about certain variables as being "greater" or "less," but precisely the same kinds of statements are required in the verbal theory.

Now let us compare Festinger's propositions with our model.

Hypothesis 1a: *The pressure on members to communicate to others in the group concerning "item x" increases*

---

[3] Read: The time rate of change in the discrepancy of opinion is a function of the pressure to communicate, the receptivity to influence, and the existing discrepancy. The remaining four equations may be read in an analogous fashion. The change in the "is a function of" designation from f to P to L to g to U is simply to indicate that the functions involved in the five equations are not necessarily identical in form.

monotonically with increase in the perceived discrepancy in opinion concerning "item x" among members of the group.

This proposition states that there is a relation between P and D, and that a change in the latter brings about a change in the former in the same direction. We have expressed this in equation 1.2 by making P dependent on D. To encompass Festinger's proposition we need to add the additional hypothesis that $P_D > 0$; where P_D symbolizes the partial derivative of P with respect to D, and designates the change in P for a unit change in D when U is held constant. If the sign of the derivative is positive (greater than zero), the relation between the variables is of the form, "if x_1 increases, x_2 increases." When the derivative is negative, the relation is the inverse.

Hypothesis 1b: *The pressure on a member to communicate to others in the group concerning "item x" increases monotonically with increase in the degree of "item x" to the functioning of the group.*

This proposition states that there is a relation between P and R and that the two vary in the same direction. Festinger also postulates (2, p. 274) that an increase in R brings about an increase in P *via* an increase in U. These hypotheses are represented by our equations 1.2 and 1.5. To translate the whole of Festinger's hypothesis we need to postulate further that $P_U > 0$ and $U_R > 0$. Note that in this instance Festinger's more formal hypothesis was supplemented by an explanatory discussion linking the more simply stated proposition to other parts of the system.

Hypothesis 1c: *The pressure on members to communicate to others in the group concerning "item x" increases monotonically with increase in the cohesiveness of the group.*

This proposition is precisely parallel to 1b, with C replacing R. Hence, it is translated by equations 1.2 and 1.5 with the conditions that $U_C > 0$ and $P_U > 0$.

Hypothesis 3a: *The amount of change in opinion resulting from receiving a communication will increase as the pressure toward uniformity in the group increases.*

The amount of change in opinion—in the direction of greater or less uniformity—between times t_0 and t_1, is expressed by the integral of dD/dt from t_0 to t_1, dD/dt being defined by equation 1.1. Let us assume, now, that $f_p < 0$ and $f_L < 0$; that is, the greater the pressure to communicate and the greater the receptivity to influence, the more rapid will be the change (decrease) in the discrepancy of opinion. We assume further, with Festinger, that $L_U > 0$ in equation 1.3. We have already assumed that $P_U > 0$. It follows that $\frac{\partial f}{\partial U} = f_L L_U + f_P P_U < 0$, and hence that the rate of change of opinion toward uniformity will be the greater, the larger is U. The total change of opinion in the interval t_0 to t_1 will be the greater, the larger is dD/dt during this interval. Represent this total change in D by ΔD, i.e.,

$$\Delta D = \int_{t_0}^{t_1} \frac{dD(t)}{dt}\, dt. \qquad [1.6]$$

Then Festinger's hypothesis 3a is equivalent to the foregoing statement that ΔD is the greater, the larger is U during the interval t_0 to t_1.

Hypothesis 3b: *The amount of change in opinion resulting from receiving a communication will increase as the strength of the resultant force to remain in the group increases for the recipient.*

This hypothesis states that ΔD is the greater, the larger is C. But C is related to U, via equation 1.5, with the requirement that $U_C > 0$. And U, in turn, is related to L via equation 1.3, with $L_U > 0$. Then, since dD/dt is related in equation 1.1 to L, with $\partial f/\partial L < 0$, it follows that the integral of dD/dt for the interval t_0 to t_1 (or ΔD) will be greater, the larger is C. This indicates that Festinger's hypothesis 3b is not independent but follows from the earlier assumptions, namely those made in 1c and 3a.

We have now established an interpretation of Festinger's five hypotheses in terms of (*a*) four equations involving six variables (we have not yet made use of equation 1.4); and (*b*) a number of statements about the signs of the partial derivatives of the dependent variables with respect to the independent variables in these equations.[4] For convenient reference, let us restate the assumptions implied by the hypotheses about the signs of partial derivatives:

$$f_P < 0, \quad [1.1a] \qquad f_L < 0, \quad [1.1b]$$

$$P_D > 0, \quad [1.2a] \qquad P_U > 0, \quad [1.2b]$$

$$L_U > 0, \quad [1.3a]$$

$$U_C > 0, \quad [1.5a] \qquad U_R > 0. \quad [1.5b]$$

Thus far we have said nothing about equation 1.4. This equation is not embodied in any of Festinger's more formal hypotheses, but is a translation of his statement that "People tend to locomote *into* [groups which share their opinions and attitudes] and *out of* groups which do not agree with them" (**2**, p. 273). Equation 1.4 states that the rate of change in cohesiveness of a group depends upon the discrepancy of opinion and the pressure toward uniformity. We postulate further that:

$$g_D < 0, \quad [1.4a] \qquad g_U < 0. \quad [1.4b]$$

That is, a large discrepancy of opinion, for a given pressure toward uniformity, or a large pressure toward uniformity, for a given discrepancy in opinion, both bring about a rapid decrease in group cohesiveness. Note that we have also postulated that the change in cohesion is dependent upon the level of cohesion itself. Equation 1.4, or some alternative, is needed to make the dynamic system complete. We shall see that in some of the empirical studies, the mechanism of equation 1.4 is in fact involved; in others of the studies we need to assume instead that C is constant for the time interval of the experiment, that is, $dC/dt = 0$.

Interpretations have now been provided for all five equations, but nothing has been said about the signs of two of the partial derivatives, f_D in equation 1.1, and g_C in equation 1.4. Let us suppose a system governed by these five equations to be in equilibrium, so that $dD/dt = 0$, and $dC/dt = 0$. Then the values of the five dependent variables, D, P, L, C, and U will depend only on the value of the single remaining independent variable R. Each value of R will determine a corresponding equilibrium position of the system. Now if P and L vary, as the system moves from one position of equilibrium to another, a determinate change in D must take place in order to satisfy equation 1.1. The direction of change in D, for increases in P and L, will depend on the signs of the two partial derivatives already assumed to be negative: f_P and f_L; and upon the sign of f_D. In this situation, using the chain rule for differentiation,[5] we get:

---

[4] Since the appearance or nonappearance of particular variables in particular equations and the signs of the partial derivatives in these equations do not depend upon the scales on which the variables are measured, our hypotheses satisfy the condition, earlier stated, that they should require only ordinal, and not cardinal, measurement of the variables.

$$f_P \delta P + f_L \delta L + f_D \delta D = 0, \qquad [1.7]$$

where δP, δL, and δD represent the changes in these variables in moving from one equilibrium position to a neighboring one. Applying the chain rule to equation 1.4 we find that:

$$g_D \delta D + g_U \delta U + g_C \delta C = 0. \quad [1.8]$$

On the basis of the interpretations we have given to the mechanisms, it is plausible to assume that if pressure to communicate (P) and receptivity (L) are large, then the discrepancy of opinion (D) will be pushed down to a lower equilibrium level through the mechanism of equation 1.1 than if P and L are small. This is equivalent to saying that $\delta D/\delta P = -(f_P/f_D) < 0$, and $\delta D/\delta L = -(f_L/f_D) < 0$; whence,

$$f_D < 0. \qquad [1.1c]$$

Similarly, high levels of discrepancy of opinion (D) and of pressure toward uniformity (U), if maintained, may be assumed to drive cohesiveness (C) to lower levels, through the mechanism of equation 1.4, than will low levels of D and U. It then follows by an argument similar to that just given that

$$g_C > 0. \qquad [1.4c]$$

This completes the comparison of Festinger's propositions with our model. Note that in addition to the five equations, it was necessary to make explicit assumptions about the signs of eleven partial derivatives. By way of summary, we diagram in Fig. 1 the causal relations among the variables that are implied by the equations. For simplicity we assume changes in the variables to take place at discrete intervals in time rather than continuously. Thus $D(t+1)$ is the value of D one time interval after t. By equation 1.1, $D\ (t+1)$ is determined by $D(t)$, $P(t)$, and $L(t)$. By equation 1.2, $P(t)$ is determined by $D(t)$ and $U(t)$; by equation 1.3, $L(t)$ is determined by $U(t)$; by equation 1.4, $C(t+1)$ is determined by $C(t)$, $D(t)$, and $U(t)$; and by equation 1.5, $U(t)$ is determined by R and $C(t)$.

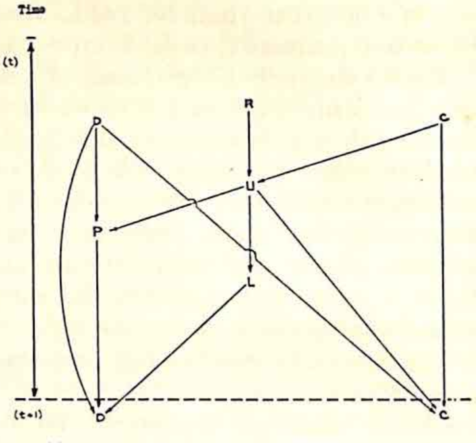


FIG. 1. Psychological mechanisms postulated in the model.

2. EXAMINATION OF THE EMPIRICAL STUDIES

Festinger draws upon evidence obtained in a number of experiments and a field study for confirmation of the hypotheses. In this section we shall re-examine the evidence from three of the studies in terms of the aggregative model that has been formulated. Our purposes are twofold: (*a*) We want to check the extent to which the empirical data support or contradict our reformulation of Festinger's system; and (*b*) we want to know whether all of the eleven "derivative" hypotheses have been tested

---

[5] See the chapter on partial differentiation in any standard calculus, e.g., Granville, *Elements of the Differential and Integral Calculus*, (rev. ed.), 1929, p. 419; or Osgood, *Advanced Calculus*, 1925, pp. 112–119.

—and if not, just which ones remain to be tested in future experimental work.

Back's Influence Experiment. Back (1) performed an experiment with teams of two subjects each. Each team member was instructed to write an interpretation of pictures he had seen alone before discussion with his teammate, and again after discussion. Each subject believed his pictures were identical with those seen by his teammate, but they were in fact somewhat different.

Opinion change in a team was measured by a content analysis of the similarities and differences in the interpretations of teammates. In this way, a measure could have been obtained of D_O, the discrepancy at the beginning of the trial, and D_T the discrepancy at the end of the trial. Back did not, in fact, measure D_O and D_T. Instead, he noticed the number of changes that took place between the initial and final interpretations in the direction of the partner's interpretation (**1**, Tables 4 and 5). The number of such changes may be taken as a measure of the *decrease* in difference in opinion, i.e., $D_O - D_T$. Since the times of all trials were short and equal, $D_O - D_T$ will be approximately proportional to dD/dt at time t_O. We shall accordingly represent it by the symbol ΔD_O.

No measurement was made that can be regarded unequivocally as a direct measure of pressure to communicate (P), receptivity (L), or pressure toward uniformity (U). This is hardly surprising, since all these variables refer to "states" of the subject, lie entirely inside his skin, and hence can be observed only indirectly through behaviors they produce. Back made three sets of observations intended to serve as indices of these latent variables. The first measure was a rating of the team discussion by observers as *active* or *withdrawing*. An active pattern will be interpreted as evidence of high P, a withdrawing pattern as evidence of low P. The second measure was the report of subjects as to whether they felt, in the discussion situation, some or no pressure from their partners to change their interpretations. Reported pressure will be interpreted as high U, lack of pressure as low U. Third, subjects were also asked to report whether they had been receptive to influence by the partner. These reports will be regarded as measures of L.

No direct measurement was made of cohesiveness or relevance. The instructions given to team members at the beginning of the trials were intended to produce high cohesiveness in half the teams and low cohesiveness in the others. The instructions, as interpreted by Back, are given as indirect measures of C_O. That certain instructions to the subjects will produce high or low cohesiveness is based only on "common sense" and introspective interpretation of the psychological impact of the instructions. Hence, it would have been better if C also had been measured directly.

It is not clear whether the intergroup differences produced by the instructions should be regarded only as differences in C_O, or whether they should be regarded as differences in R_O as well. In three different sets of experimental groups, Back attempted to employ three different bases for inducing high or low cohesiveness: personal attraction to the team partner, reward for group performance, and the prestige of the group. Back argues that pressure toward uniformity arises from two sources: the function of the group as a reference group in determining social reality (which depends on *cohesiveness*), and the function of the group as a means toward reaching a personal goal (which operates through the *relevance* of uniformity of opinion to goal achievement). It is reasonable, therefore, to interpret the first and third of the motivations in Back's experiment as cohesiveness-producing motivations; but to interpret the second—reward for performance—as a relevance-producing motivation. Hence, we shall assume that the experimental instructions produce differences both in C_O and in R_O for all three sets of groups. In the present interpretations of the model this distinction is not vital, for U_C and U_R are both positive, and hence act upon U in equation 1.5 in the same direction. A clarification of the distinction would appear important, however, to further development and testing of the theoretical scheme, if any operational significance is to be attached to the asserted difference between these two variables.

It is implicitly assumed in the experimental design involving team discussions of about ten minutes that the effect of the instructions upon cohesiveness and relevance persisted for the duration of the experiment. If this is correct we can eliminate equation 1.4 from the system, and assume that C, like R, in equation 1.5 is a parameter, constant for the duration of the experiment with any one group. This amounts to an interpretation of this experiment in terms of a "short-run" model in which the mechanisms represented by the remaining four equations were operating, but in which the system did not have time to settle to a final equilibrium. We simply ignore the "feedback" effect of dD/dt on D. Then we may regard the equations as four relations that determine the values of the dependent variables ΔD_O, P_O, L_O, and U_O as functions of the independent variables (or "parameters") D_O, C_O, and R_O. Finally, since D_O is assumed to be the same for all groups (the same sets of pictures were shown to all groups), this variable may be ignored in analyzing the intergroup differences.

What predictions can we make from the model of the Back experiment so interpreted? If we employ symbols like δP, again, to designate intergroup differences in the initial values of the dependent variables as functions of intergroup differences in the independent variables, we can deduce the following from our previous equations by differentiating them, using the chain rule:

$$\delta U_O = U_C \delta C + U_R \delta R, \quad \text{(from equation 1.5)}, \quad [2.1]$$

$$\delta L_O = L_U \delta U, \quad \text{(from equation 1.3)}, \quad [2.2]$$

$$\delta P_O = P_U \delta U, \quad \text{(from equation 1.2)}, \quad [2.3]$$

$$\delta \Delta D_O = f_P \delta P + f_L \delta L, \quad \text{(from equation 1.1)}. \quad [2.4]$$

From these relations, together with the previous assumptions as to the signs of partial derivatives, the following hypotheses can be derived by algebraic means, taking C and R as independent variables:

$$\delta U/\delta C = U_C > 0, \quad [2.1a]$$

$$\delta U/\delta R = U_R > 0, \quad [2.1b]$$

$$\delta L/\delta C = L_U U_C > 0, \quad [2.2a]$$

$$\delta L/\delta R = L_U U_R > 0, \quad [2.2b]$$

$$\delta P/\delta C = P_U U_C > 0, \quad [2.3a]$$

$$\delta P/\delta R = P_U U_R > 0, \quad [2.3b]$$

$$\delta \Delta D/\delta C = (f_P P_U + f_L L_U) U_C < 0, \quad [2.4a]$$

$$\delta \Delta D/\delta R = (f_P P_U + f_L L_U) U_R < 0. \quad [2.4b]$$

Let us now check the extent to which Back's empirical findings confirm these predictions. The fact that persons in the high cohesive groups tried to influence (P) their partners more actively (chi square significant at .02 level) than did individuals in the low cohesive groups, who tended to withdraw (**1**, p. 15), may be taken as confirmation of equations 2.3a and 2.3b. His finding that "less than half of the members of the low cohesive group reported that they felt some pressure (U), while more than two thirds of the members of the high cohesive group did so" (**1** p. 16) (again significant at the .02 level), does not contradict equations 2.1a and 2.1b. Although Back says "Self-ratings on resistance [the inverse of L] . . . show a slight decrease" (**1**, p. 16) in level in the high cohesive groups, he reports the difference is not statistically significant. Thus, there is only weak confirmation of equations 2.2a or 2.2b.

Table 4 in Back's report (**1**) provides confirmation for equation 2.4a, $\delta\Delta D/\delta C < 0$; and equation 2.4b, $\delta\Delta D/\delta R < 0$. These significant findings are *consistent* with the assumptions of the original model but *do not*, it should be noted, *verify* all the individual assumptions about the signs of partial derivatives. The lack of confirmation, at a significant level, of equations 2.2a and 2.2b is disturbing, since $L_U U_C$ or $L_U U_R$, or both, appear in all the other hypotheses, too. Since $\partial L/\partial C$ and $\partial L/\partial R$ are found empirically not to be significantly different from zero, either L_U or both U_C and U_R must be close to zero. But inasmuch as confirmation of equations 2.1a and 2.1b indicates that $U_C \neq 0$, L_U must be close to zero; such a conclusion is not contradicted by our confirmation of equations 2.4a and 2.4b, for the $f_L L_U$ terms might be zero without changing the signs of $\delta\Delta D/\delta C$ or $\delta\Delta D/\delta R$.

In summary, we may say that Back's experiment constitutes a test of some short-run properties of the present mathematical model. The findings of the experiment verified some, but not all, of the assumptions regarding the signs of partial derivatives, and strongly contradicted none of the assumptions. If the experiment were to be repeated, its power to test the model could be increased by introducing additional procedures for measuring certain of the variables, such as R, C_O, and C_T.

The Festinger-Thibaut Experiment on Interpersonal Communication. This experiment (**4**) involved variables referring to individual members of a group, as well as variables defined as group averages. The present discussion is limited to those findings of the experiment that can be stated in terms of aggregates.

A problem was given to members of a discussion group, and they were asked to record their solutions before and after discussion. The standard deviations of the opinions in each group at the beginning of the twenty-minute discussion and at the end are taken as measurements of the variable D—specifically, D_O and D_T. The several experimental groups were initially given different sets of instructions. The instructions that aimed at inducing "homogeneity" stressed the similarity of group members, and can be interpreted as cohesiveness-producing variables. The instructions aimed at producing "pressure toward uniformity" stated that the group should arrive at agreement on opinion or at the "correct" opinion. In Festinger's scheme, this variable may be translated as *relevance*. We shall treat the homogeneity-heterogeneity instructions as determining C_O, and the other instructions as determining R_O.

The appropriate model for this experiment would appear to be the same as for the Back experiment. The experimental findings can be expressed (**4**, p. 99, Fig. 2) as $\delta \Delta D/\delta C < 0$ and $\delta \Delta D/\delta R < 0$. Thus we see that the findings in this experiment exactly parallel one of the findings in the Back experiment.

The need for further clarification of the variables C and R is evidenced by the fact that Festinger cites Back's experiment as evidence for his hypothesis 3b, and the Festinger-Thibaut study as evidence for his hypothesis 3a, while we have concluded that there is no difference in the findings regarding ΔD in the two experiments. The reason for this difference is that Festinger treats the instructions in the Back experiment as determining C and instructions in the Festinger-Thibaut experiment as determining U, while we have interpreted these instructions as determining C and R in both experiments.

The Housing Study. Unlike the two experiments already discussed, the housing study (**3**) was a field study and not an experiment. This has two consequences. First, the field situation reflects the working of *all* the mechanisms actually present. Individual variables cannot be made "independent" through experimental control and randomization, as they can in the laboratory. Second, the relevant time span during which the mechanisms operate is much longer in the field situation than in the laboratory. In a relatively stable field situation of this sort we can expect to observe the variables only in the neighborhood of the equilibrium position or positions of the system.

In the field study, a correlation of .74 was found between (*a*) the cohesiveness of groups in a housing project, and (*b*) the effectiveness with which a group standard relevant to the functioning of the groups was maintained (**3**, p. 12, Table 16). Festinger treats these findings as parallel to the findings in Back's work in confirmation of hypothesis 3b.

In interpreting the earlier experiments, because of the short time period involved, the feedback described by equations 1.1 and 1.4 could be ignored. Moreover, the variable D_O was assumed to be the same for all groups at the outset of the experiment. In the field study, there is no reason to suppose that we can ignore the change of D and C through time, as determined by equations 1.1 and 1.4. Furthermore, the variables that are correlated are D_T with C_T, and not D_O with C_O and R_O, for the measurements were made after the residents of the housing project had been residing there and interacting for a considerable period (up to 15 months).

These considerations suggest that the appropriate model for representing the field study is the complete model of equations 1.1 through 1.5, together with the corresponding inequalities. If this is so, then the short-run assumptions that permitted us to apply the method of comparative statics in the two experiments are no longer applicable. Given the initial conditions, the time path of the behavior of the group will be determined by equations 1.1 through 1.5, and two groups with the same initial positions will have the same paths. If we assume that the initial values of at least some variables were different for the various groups, we can derive a plausible interpretation of the field study, and can show also that the field results prove something quite different from (but not inconsistent with) Back's experiment—contrary to Festinger's interpretation of the two sets of data as supporting the same proposition.

Using equations 1.2, 1.3, and 1.5 to solve for P, L, and U, respectively, the system can be rewritten in terms of two different equations in C and D:

$$\frac{dD}{dt} = \mathrm{f}\{P[D, U(C)], L[U(C)], D\}, \qquad [2.6]$$

$$\frac{dC}{dt} = \mathrm{g}\{D, U(C), C\}. \qquad [2.7]$$

Now for a system obeying these equations and starting off from some initial position $[D_O, C_O]$ we can represent its path by a curve in the D-C plane. Each point on the curve $[D(t), C(t)]$ will represent the position of the system at some time, t. What can we say in general about the various paths corresponding to different initial conditions?[6]

Consider the points for which $dD/dt = f = 0$. These lie on a curve on the D-C plane. Moving along this curve, the requirement that $f = 0$ gives us the following equation for the ratio between the rate of change of C, i.e., δC, and the rate of change of D, δD:

$$f_P\{P_D\delta D + P_U U_C \delta C\} + f_L L_U U_C \delta C + f_D \delta D = 0. \qquad [2.8]$$

This expression is obtained by differentiating equation 2.7, using the chain rule. Then we have for the slope of this curve:

$$\rho = \left[\frac{\delta C}{\delta D}\right]_{dD/dt=0} = -\frac{(f_P P_D + f_D)}{U_C(f_P P_U + f_L L_U)} < 0. \qquad [2.9]$$

Similarly, for the curve on which $dC/dt = g = 0$, we obtain:

$$\sigma = \left[\frac{\delta C}{\delta D}\right]_{dC/dt=0} = -\frac{g_D}{g_U U_C + g_C} < 0. \qquad [2.10]$$

By using our assumptions about the signs of these partial derivatives already made above, it can be shown that the slopes of both these curves are negative in the D-C plane.

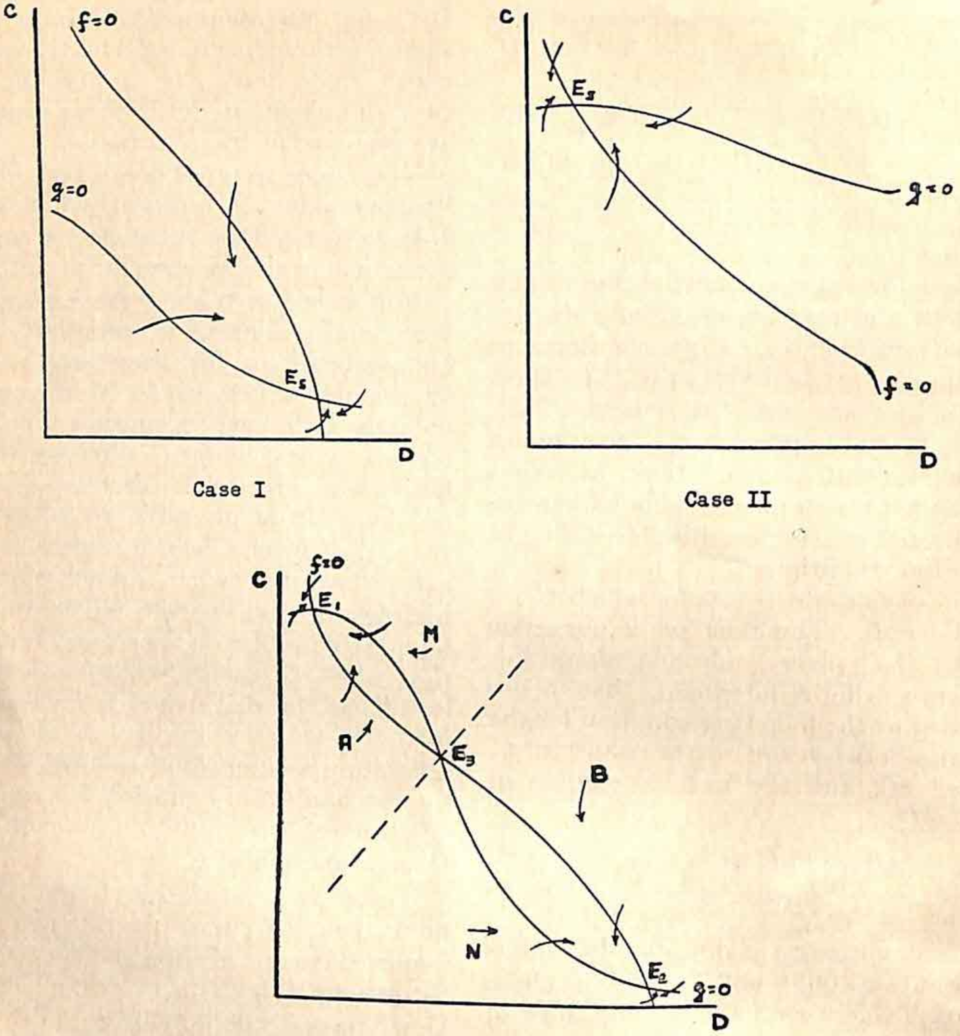
Now there are grounds of plausibility for making assumptions which enable us to say that these curves have a particular shape—specifically, that σ approaches zero for very large and very small values of C, and that $1/\rho$ approaches zero for very large and very small values of D. The argument is as follows: Equation 2.7 describes a mechanism whereby cohesiveness adjusts to discrepancy of opinion. It is plausible to suppose that this mechanism is subject to saturation—that when C reaches very high levels it will not be driven much higher by further decreases in D; and that when C reaches very low levels it will not be driven much lower by further increases in D. A similar saturation assumption is plausible for the mechanism of equation 2.6 which determines the equilibrium level of D as a function of C.[7]

If the two saturation hypotheses hold, then the curves $f = 0$ and $g = 0$ must have one of three possible configurations (Fig. 2). In cases I and II we have a single point of equilibrium—a point for which simultaneously $f = dD/dt = 0$, and $g = dC/dt = 0$. In case III we have three such points of equilibrium. An analysis by standard methods (omitted here) shows that certain conditions must be satisfied if the equilibrium position is

---

[6] For the methods applied here, see Lester R. Ford, *Differential Equations*, pp. 9–12; and A. A. Andronow and C. E. Chaikin, *Theory of Oscillations*, 1949, pp. 8–12, 182–193, 203–8.

[7] These two saturation hypotheses each imply cardinal measurement of C and D. The conclusions we shall draw from the two of them together, with respect to the existence and location of stable and unstable equilibrium points, depend, however, on only ordinal measurement. It would be desirable to restate the saturation hypotheses—or that part of them we need—so as to remove the cardinality assumptions. This must be possible, but we have not yet been able to do it.



Case I

Case II

Case III

FIG. 2. Single and multiple equilibriums.

to be dynamically stable (so that if the system is disturbed slightly from equilibrium, it will tend to move back toward the equilibrium position). The stability condition is:

$$\sigma/\rho < 1. \qquad [2.11]$$

We observe from Fig. 2 that equation 2.11 is satisfied for the equilibrium points (E_s) in case I and case II, and for the two extreme equilibrium points (E_1 and E_2) in case III.

In the housing study tenants were assigned to particular courts in the housing project, virtually at random. The initial cohesiveness ("tendency to associate") of residents of a court might depend largely on the opportunities people found to meet others inside and outside the project. The initial discrepancy of opinion in a court with regard to a tenant's organization might be regarded as a chance variable depending on the past experiences of the tenants with such neighborhood activities. Quite by chance, some courts would have a higher initial D or C than others.

Now suppose that two groups (A and B) initially had the same cohesiveness within their courts, but that there was much less initial divergence of opinions with regard to the desirability of a tenants' organization in one (A) than in the other (B), as is illustrated in case III of Fig. 3. If the D_0 of a particular court is low, as in A, the cohesiveness of the tenants will increase, which in turn will cause them to accommodate their opinions to each others', further decreasing D. If the D_0 of a court is high, as in B, the cohesiveness of the tenants will decrease, which will promote further divergence of opinion. All these consequences follow from the postulates of the mathematical model. Hence, when the system reaches equilibrium we should expect to find some courts with high C and low D, and some with low C and high D, giving a negative correlation between C and D. Further, as the process approaches equilibrium, one would expect a clustering of points (representing the courts) around the two stable equilibria, E_1 and E_2. A plot of the data from the Festinger, Schachter, and Back study, as presented in Fig. 3, case III, reveals just such a clustering, giving empirical confirmation of our postulation that the feedback process has been in operation.

This interpretation of the field work permits a prediction of the observed correlation from the postulates of the mathematical model. The predicted correlation depends, however, on a larger set of hypotheses than those used in the interpretation of the Back and the Festinger-Thibaut experiments. To account for the findings, both feedbacks provided in equations 1.1 and 1.4 were needed, and the two saturation assumptions were added. Festinger's original interpretation of the laboratory and field results as supporting hypothesis 3b was hardly accurate. Our formalization of the hypothesis shows its complexity, and indicates that the Back experiment supports the short-run static model, while the Festinger, Schachter, and Back data support the enlarged dynamic model.

Note that only case III of Fig. 2 predicts a negative correlation; cases I and II predict that the points in the C-D plane should cluster around E_3 thus yielding a zero correlation. Yet data supporting any one of the three cases would be consonant with the

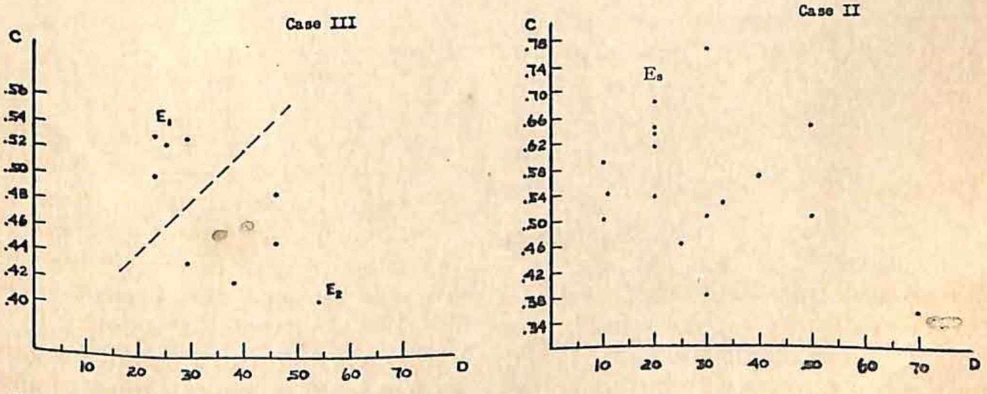


FIG. 3. Empirical tests of the equilibrium cases. In case III, data are from reference 3, p. 92, Table 16; in case II, data are from reference 3, p. 93, Table 17.

hypotheses used in developing the model. Which of the three cases will hold depends upon the equilibrium equations (the relative positions of the two equilibrium curves on the D-C plane). If we plot the data obtained in the second housing project (**3**, p. 93, Table 17), we find that these data seem to correspond to case II, presented in the "Case II" graph in Fig. 3. According to our arguments, Festinger, Schachter, and Back incorrectly interpret their zero correlation as indicating no feedback had yet occurred in the few months Westgate West had been in operation. The clustering of points at E_s indicates that the dynamic model is applicable, and that some groups in this cluster of points are close to equilibrium. The zero correlation between C_T and D_T actually confirms the existence of feedback.

SUMMARY

In this paper we have undertaken to carry further the development of a set of hypotheses about pressures toward uniformity in groups into a more integrated, less redundant system. We have constructed explicit postulates about the intervening mechanisms to provide linkages among variables, such as "Relevance" and "Pressure to Communicate." Although the model covers static situations in which there is no feedback, it also postulates dynamic mechanisms by which, for example, cohesion at one time influences the cohesion of the same group at a later time. In the present paper, all the variables considered were averages of the behaviors of all members of a group.

A number of deductions were obtained from the system. Some of these were verified by bringing field and experimental data to bear. The model led us to examine correlational data for bimodality characteristics, indicating the existence of two stable equilibrium points when groups experience uniformity.

REFERENCES

1. BACK, K. W. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, 46, 9–23.
2. FESTINGER, L. Informal social communication. *Psychol. Rev.*, 1950, **57** 271–282.
3. FESTINGER, L., SCHACHTER, S., & BACK, K. *Social pressures in informal groups.* New York: Harper, 1950.
4. FESTINGER, L., & THIBAUT, J. Interpersonal communication in small groups. *J. abnorm. soc. Psychol.*, 1951, **46**, 92–99.

(Received February 16, 1954)

Psychological Review
Vol. 62, No. 1, 1955

THE CONCEPT OF EFFICIENCY IN PSYCHOLOGICAL HEALTH AND IN PSYCHOPATHOLOGY[1]

JULIUS WISHNER

University of Pennsylvania

The central problem of this paper is to formulate a descriptive concept of psychological health and psychopathology. We shall be concerned more with the adequacy of the conception for the laboratory situation than for its immediate applicability to clinical formulations or to psychotherapy.

By way of introduction, it might be useful to point to certain differences in the implications of the terms "health" and "normality." It is true, of course, that the two are often used interchangeably, but this leads to confusion in conception and is a reflection of the multiple meanings, sometimes mutually exclusive, attached to the terms (cf. 9). We may distinguish between them by using "normal" only in relation to the normal curve of distribution. Abnormality, then, represents deviation from some norm. It does not, in itself, convey information concerning the adequacy or appropriateness of behavior. Health, on the other hand, implies precisely such information; it does not necessarily convey information concerning norms or frequency of occurrence of some phenomenon. Understood in this way, the discussion to follow will be concerned with the problem of psychological health and psychopathology rather than with normality and abnormality. Nevertheless, experimentation utilizing both kinds of conceptions will be cited, since there has been no general agreement on a standardized usage. It should be noted that to eschew a wholly relativistic notion of health in terms of normality does not mean that the criteria must necessarily be of the absolutistic variety. The object is to formulate a dimension in which any two behavior patterns can be compared as to the degree of health indicated. The behavior need be referred neither to an absolute, a priori, criterion nor to some empirical norm.

Further, it should be seen that the formulation of such a dimension is not necessarily related to the social definition of psychological health. This may vary from society to society and from time to time, presumably depending upon social conditions, degree of knowledge, availability of treatment facilities, etc. However, what a particular society says is healthy at a given time is not necessarily central to a scientific formulation of a dimension of psychological health. A particular society may believe malaria to be a sign of health, but the definition of physical health need not be concerned with such an opinion. Certainly such opinions can become important questions for many kinds of research, as for example, in investigations of attitudes, but they are not of central significance in the formulation of an objective definition of the field.

PSYCHOLOGICAL HEALTH AS EFFICIENCY IN MEETING ENVIRONMENTAL REQUIREMENTS; PSYCHOPHYSIOLOGICAL AND PSYCHOMOTOR EXPERIMENTS

Individual cases of psychopathology as seen in the clinic present the investigator with such a multiplicity of different problems that it does not appear likely that they can all be subsumed under a single rubric with respect to their origin. Nevertheless, they are recognized as having in common some form of psychopathology. Although various classificatory systems have been offered which attempt to categorize such individuals into classes differing in kind, there remains a basic assumption that they can also be ordered with respect to the degree of psychopathology shown; i.e., individual cases are recognized as being more or less disturbed.

---

[1] Based on a paper presented at a symposium on "The Meaning of Psychological Health" at the Eastern Psychological Association meetings in Boston, April 1953. The writer is indebted to Mr. Thomas E. Shipley, Jr., for aid in the formulation, and to several colleagues for critical comments.

What are the criteria of disturbance? Many have been proposed. In general, they can be grouped into two classes, which are not necessarily mutually exclusive: (*a*) those emphasizing subjective feelings, personal discomfort, persistent feelings of tension, etc., and (*b*) those emphasizing "appropriateness" or "adequacy" of objectively measurable behavior. For the most part, the textbooks try to combine these (e.g., 4). Without entering into the endless controversy over the possibility of the scientific study of subjective feelings, the present discussion will be confined to a consideration of objectively measurable behavior.

On the basis of experimental data, the argument will be developed that the continuum of psychological health and psychopathology can be conceptualized in terms of the degree of efficiency in meeting environmental requirements. The definition of efficiency and environmental requirements is one of the tasks of this section, and will be discussed presently. For the moment, it is well to indicate that inefficiency may be focalized at the perceptual, cognitive, feeling, or action levels, and the seriousness of the consequences may be a function of the particular focus. The essential point is, however, that without some inefficiency no psychopathology would be recognized.

The formulation proposed here has as its principal point of departure those definitions of psychological health emphasizing appropriateness and adequacy of behavior. Aside from the more specific operational analysis to be given for the present terminology, the term efficiency is preferred here to such terms as appropriateness or adequacy because the latter usually involve a priori value judgments, which would make an objective definition of psychological health impossible. Moreover, efficiency immediately implies something quantifiable, whereas appropriateness and adequacy are usually considered in qualitative terms. Apart from this, all three terms, as used in this paper, have similar connotations.

Efficiency is a meaningful parameter only when the most important aspects of the task to be completed can be specified. Thus, with respect to behavior in "everyday life," the difficulties encountered in the use of such concepts as appropriateness are not surmounted by the use of the term efficiency. For the essential stumbling block remains: our inability to specify objectively what behavior is required. This is neither surprising nor crucial. The everyday situation is experimentally uncontrollable, and the testing of scientific hypotheses requires controlled conditions. A broken bone may not be diagnosable in everyday life. The fact that an X ray must be taken under controlled conditions makes it no less relevant to general anatomical integrity. Similarly, efficiency may not be measurable in everyday life, but this does not affect its relevance to the problem of psychological health if a relationship can be shown to hold under experimentally controlled conditions. With this argument in mind, we can turn to a consideration of the methods of measurement of efficiency in the psychological laboratory, and their relationship to the continuum of psychopathology.

The laboratory situation provides a ready method for specifying the task requirements. These consist of the experimenter's instructions and the relation of these to the stimulating field. If the subjects are motivated to cooperate with the experimenter, the task requirements can be held relatively constant, but their structure, i.e., their "requiredness-qualities" must be known. Thus, the problem for controlled experimentation should be centered on finding a method for measuring efficiency and the clear specification of task requirements.

In psychological experiments, efficiency has often been equated with productivity. It should be apparent immediately that such a measure cannot be very useful in relation to psychological health and psychopathology, for it takes no account of the comparative effort exerted to accomplish a given task. It is a common observation that neurotics will often achieve some end, but at a terrible price to themselves, sometimes to the point of breakdown. Productivity, as measured by speed or complexity of performance, is indeed an important dimension of performance, but it is not, by itself, an index of efficiency. The latter requires a more complex analysis.

In inanimate machines, efficiency is a function of the ratio of output to input, and the latter can be controlled while the former is measured. In animals, however, input, in the same sense as in a machine, is *an unknown quantity in relation to any given task*. Only the direction of energy expenditure can be observed. In a previous paper (24), it was suggested that, with respect to any given task requirements, at least two components of the energy expended can be distinguished: (*a*) focused behavior (*F*), energy expended directly on the task, and (*b*) diffuse behavior (*D*), energy expended in irrelevant directions. The most common method of measurement consists in recording simultaneous indices of tension, such as muscle action potentials, from a part of the body immediately involved in work and a part of the body not thus involved. A mirror-tracing experiment by Telford and Swenson (20) may serve as an illustration. They recorded pressures exerted on bulbs by both hands while one was occupied with the tracing. As performance improved, they found a decrease in pressure in the unused hand and an increase in the used hand. Such results support the notion that a comparison of simultaneous *D* and *F* activity may be a valid index of efficiency.

In the paper previously cited (24), evidence was adduced for the hypothesis that individuals with psychopathology are characterized by a preponderance of *D* activity, as compared to *F*. Other findings point in the same direction, although, as will be seen, an additional factor probably needs to be introduced because of the inadequacy of our present analyses of tasks and of our measuring techniques. Luria (10) was concerned with the relationship between psychic conflicts and motor as well as associative processes. He used the word association technique, and his subjects were instructed to press down on a tambour with their right hand as soon as they gave a response word to the stimulus. Measurements were made of the movement of the left hand as well as of the right hand. Luria found that some form of motor disorganization and verbal blocking accompanied psychological disturbance, whether induced or existing in its natural state. Similar findings have been reported by Duffy (6) and Arnold (1).

An analysis of Luria's task might be useful here since much of the current work in this area utilizes his technique as a point of departure. In terms of *D* and *F* activity, it would appear that the performance of the instructed right-hand movement in close coordination with the verbalization of the response word would constitute *F* activity; other movements of the right hand and all movements of the left hand would constitute *D* activity. This analysis leaves something to be desired, however. For example, there is no meaningful way of quantifying *F* activity. The intensity of right-hand pressure might represent either *D* or *F* activity. Pushing down hard on the tambour with the right hand could be a reflection of intense disturbance which might not be seen in any other aspect of motor activity measured. The essential difficulty lies in the fact that the activity called for is not an integral part of the whole task, the central requirement of which is associating a meaningful word to the stimulus as rapidly as possible. Much cognitive activity occurs in this process, and there is no way of telling which part of it is *D* and which *F*. The required movement of the right hand has been appended arbitrarily by the experimenter to the central need for a word association; i.e., the structure of the situation demands intrinsically a verbal response. The motor activity requested is essentially ancillary. That it is an arbitrary appendage to the central structure of the task can be seen if it is recognized that it can just as easily be included in a host of other situations. This makes Luria's findings all the more interesting, since, despite the lack of inherent relationships of the movements to the task, psychic conflict is still reflected in *D* activity.

The same critique can be applied to a recent series of rather ingenious experiments by Malmo *et al.*, only two of which will be discussed here (11, 12). In one experiment (11), these investigators utilized the Luria technique with electronic measurement, with four groups of subjects: controls, psychoneurotics, acute psychotics, and chronic schizophrenics. The degree of motor disorganization found in the various groups is just what might have been expected on the basis of clinical experience. The chronic schizophrenics were most disorganized, the acute psychotics were next, the psychoneurotics next, and the controls were significantly different from all other groups. At the same time, there were no significant differences between the controls, psychoneurotics, and acute psychotics in productivity, as measured by the number of accurate discriminations of sizes of circles. Only the chronic schizophrenics were significantly different from the rest of the subjects in this respect, probably as a result of inadequate attention.

In another experiment by these investigators (12), the question of the relevance of the activity measured to the task requirements is raised even more acutely. With the same groups of subjects, they measured the changes in muscle action potentials in the right arm and neck in response to various intensities of heat stimulation to the forehead. The subjects were instructed to press down on a key as soon as they felt that the stimulus was about to become painful. It should be apparent that this situation allows for no objective specification of the task requirements. Many different kinds of evaluations of the situation by the subject are plausible, and these might well influence the nature of their responses. Nevertheless, the results are of interest here because they suggest the need for inclusion of some measure of productivity in the quantification of efficiency in the context of this discussion. If we assume, for the sake of further evaluation, that right-hand tension in this experiment represents *F* activity and that neck tension represents D activity, the findings show no difference in efficiency between chronic schizophrenics and controls. This is obviously a function of the fact that the former showed very little change in muscle tension anywhere; i.e., the chronic schizophrenics were simply unreactive to the stimulation. Qualitatively, this might be said to be inappropriate behavior. Nevertheless, if the assumptions concerning D and F activity in this experiment are tenable, the proposed *quantification* would break down in relation to a continuum of psychopathology if Malmo's findings concerning chronic schizophrenics are confirmed. By taking into account the productivity (P) of the individual, how he *actually affects his environment,* this situation can be remedied. It is known, as is shown in the first experiment cited (11), that chronic schizophrenics are underproductive. Thus, when they are not reactive, their failure to produce, if used as a multiplicative factor, would produce a low efficiency (E) rating. Thus, a general equation for efficiency in motor activity might be written: $E = f(F/D, P)$. Of course, the specification of the exact nature of the function must await empirical findings from experiments oriented to this problem. In view of the previous results, it seems likely that such an equation would reflect degrees of psychopathology in experiments oriented to this problem.

A final comment about the possibilities of measuring ostensive D and F activity in clearly defined situations seems necessary in view of the above criticism of previous experiments. What is required is that the activity be an intrinsic aspect of the total structure of the task. In addition, unless D and F activity can be clearly defined in the particular situation, complex cognitive activity will have to be circumvented, at least in psychomotor experiments; for it is entirely unclear which motor behavior is relevant to which cognitive operation. These prerequisites to the measurement of psychomotor or psychophysiological efficiency may be met by utilizing relatively simple situations, such as simple reaction time. They can also be met in more complex situations by appropriate analysis and measurement, as is seen in an old experiment by Morgan (13). He was interested in the effect of noise on the work output of typists. The subjects were required to decode a letter which the experimenter showed them and press the appropriate key on a keyboard. Measurement was made of the pressure on the key, as well as of total work output under quiet and noisy conditions. The results are interesting and point up the previous analysis of the difficulty of specifying the F activity in Luria's experiments. The effect of noise on work output was insignificant. As a matter of fact, after an initial decrease, output increased slightly as adaptation occurred. At the same time, there was an increase in pressure exerted on the keys under conditions of noise. This result also points up the probable need for the inclusion of a productivity measure in the rating of efficiency. For if increase in pressure were relatively uniform among the subjects, then it would still be possible to rate them as to their productivity, which would certainly reflect, all other things being equal, the efficiency with which they adapted to the noise situation.

To summarize this section, it has been proposed that psychological health and psychopathology be conceived as a continuum and defined in terms of the efficiency with which environmental requirements are met. It was pointed out that for the present this definition can be useful only in a relatively narrow laboratory situation because of the difficulties involved in the specification of objective requirements in the social sphere. Consideration of several experiments indicated that efficiency is a function of the ratio of focused to diffuse activity and of productivity, and that this is related to degree of psychopathology.

Thus far we have dealt with relatively simple psychomotor experiments. Attention will now be focused on efficiency in other laboratory situations.

EFFICIENCY IN LEARNING AS RELATED TO PSYCHOPATHOLOGY

Whatever may be the validity of efficiency in psychomotor activity as an index of psychopathology, it is, at best, a superficial descriptive parameter, another method of measurement. If the present proposal is to have any general significance, it will be necessary to explore its applicability to some of the more crucial fields of modern psychology. For psychopathology, these include, at least, learning and motivation, which will occupy our attention in this and the following section.

The relationship of anxiety and conditioning has occupied the attention of numerous investigators recently. It appears to be a most stable finding that in a classical conditioning situation the rate of conditioning is greater in anxious than in nonanxious individuals (3, 15, 16, 17, 18, 21, 22). The opposite findings are reported in serial learning (7, 8, 19), but apparently the nature of the task is an important variable which can affect the results markedly (14). The measures of anxiety differ from study to study, and it is still an open question whether the same independent variable is being used in each case, as Child (5) has pointed out. On the other hand, it is probable that the rate of conditioning in a classical situation is related to degree of psychopathology. Welch and Kubis (21, 22) found a significant difference between anxiety patients and normals in rate of conditioning of the galvanic skin response (GSR). Spence and Taylor (17) found a significant difference between psychotics and neurotics or normals in percentage of conditioned eyeblink responses. They report no significant difference between neurotics and normals. It is thus questionable whether anxiety, clinically identified, is a necessary condition for the reported phenomena in this area, for it is by no means clear that psychotics are always characterized by anxiety. Spence and Taylor do not state the clinical characteristics of their group. It must be noted, however, that, in the clinical situation, psychosis is often regarded as a retreat from realistic problems that give rise to anxiety.[2]

If these data are to be analyzed in terms of the degree of efficiency shown by the subjects, a thorough examination of the task requirements must be made. For this purpose the experiment by Welch and Kubis (22) is particularly well suited. They utilized two groups of subjects, 22 normals and 24 patients who presented a clinical picture of anxiety. They conditioned the galvanic skin response to the syllable "kax" which was one of 54 nonsense syllables presented to the subjects in random order. The unconditioned stimulus was a buzzer which was sounded on alternate presentations of "kax." The results showed that the anxiety patients conditioned about three times as quickly as the normals and extinguished about twice as slowly; i.e., the anxiety patients took one-third as many trials to condition as the normals and twice as many trials as the normals to extinguish. Superficial consideration of these results would seem to indicate that the anxiety patients met the requirements of the situation more efficiently than the normals. However, closer examination of the instructions to the subjects casts considerable doubt on such a conclusion. The subjects were told that it was a test of their sweating activity, that the experimenter wanted them to relax as much as possible, and that in order to help them to relax and keep their minds off extraneous matters, they would be shown a group of nonsense syllables which they were to read aloud. It is true that the experimenters wanted to force the subjects to make an arbitrary[3] association between two events contiguous in time: the nonsense syllable "kax" and the buzzer. But the subjects were not told this, and it is not stated that any of them discovered it. Now, if we assume that the two groups were equally motivated to cooperate with the experimenter, it becomes apparent that the normals fulfilled the requirements of the task, as, presumably, they saw them most efficiently. For them, the buzzer was extraneous stimulation which, according to the experimenter's instructions, they were supposed to ignore.

---

[2] The writer is aware of the many other studies in this area. Many of them have been reviewed by Child (5). They are not cited here specifically because: (*a*) of the need for economy of space; (*b*) they are not centrally germane; and (*c*) the method of analysis proposed would be essentially the same as in the studies cited.

The fact that the normals eventually were also conditioned gives rise to some interesting speculation. Is it possible that the normals were conditioned only after they became bored with the task, or sufficiently disturbed by the repeated soundings of the buzzer? Such a possibility should be given serious consideration. It seems reasonable to suppose that the anxiety patients were generally in a more diffuse state, less able to concentrate. The assumption that, in such a state, an individual is more prone to make *arbitrary* associations between events contiguous in time and space is tenable on the basis of the experimental data available. The normals condition too, when the monotony of the task disturbs their integrated functioning sufficiently to make them "victims" of contiguity, arbitrarily arranged by the experimenter. Some of the data of Taylor (18) and Spence and Taylor (16) appear to support this notion. They have plotted the mean number of conditioned eyeblink responses of their anxious and nonanxious subjects as a function of succeeding blocks of 10 trials. Taylor's curves show that the nonanxious subjects give practically no CR's for the first 20 trials, and only after 60 trials does the slope of the nonanxious curve begin to approximate the slope of the anxious curve. The Spence and Taylor curves are similar although not quite so spectacular.

There is, of course, a substantial psychological difference between being an unwitting "victim" of contiguity and *learning* by contiguity, which is often a task requirement, as in serial learning. There is a difference between memorizing something, even a list of nonsense syllables, for some purpose, as against making spontaneous, arbitrary associations between contiguous events which have no inherent connection with each other. The latter is reminiscent of the conditioning apparently involved in phobia formation.

The question arises whether the task requirements in the Welch and Kubis experiment could be changed by a change in the instructions, and so effect a reversal of the results. If the subjects were instructed that the situation contains a problem of contiguity, such a reversal might well be brought about. In other words, it would be predicted, on the basis of the present analysis, that the normals would "solve" the problem sooner, and it could be expected that this would be reflected in more rapid GSR "conditioning." Under these conditions, of course, the experiment is no longer of the classical type. It might be characterized more accurately as a problem-solving situation.[4]

---

[3] Arbitrary because there is no inherent or logical connection between "kax" and the buzzer.

In sum, then, it has been proposed that efficiency in learning can be measured if the task requirements can be specified. Efficiency in learning is characterized by a tendency to seek for orderly relationships (except when the subject is instructed otherwise), and to resist arbitrary associations. These can be induced most easily when the organism is in a diffuse state, either because of the existence of some degree of psychopathology, or because of the inherent senselessness of the task.

Efficiency in Motivation: Task Orientation and Interpersonal Cooperativeness

The purpose of this section is to adumbrate possible developments in this area, rather than attempt to deal exhaustively with it. Utilizing the concept of efficiency as defined above, the initial approach must be toward a specification of environmental requirements. Aside from the so-called primary drives, however, such specification for the field of motivation constitutes a gigantic problem beyond the scope of this paper. Nevertheless, at least two very general requirements can be assumed, although they do not exhaust the possibilities by any means. As Wertheimer (23) and Asch (2) have pointed out, social existence and creative achievement require a certain degree of interpersonal cooperativeness and task involvement respectively. Such assumptions may run counter to what Asch has labelled the "'private profit' theory of society," but they seem entirely tenable, indeed necessary, in view of the facts of social life and the phenomenology of productive thinking.

Granted these general environmental requirements, what are the implications for experimental work in psychopathology? The proposed concept of efficiency would necessitate the examination or control of the subjects' motives with respect to these requirements. The degree of concordance of the motives with the task requirements would presumably determine the efficiency of the relevant behavior. Unfortunately, published experimental data centered on this problem are unavailable. The general findings that ego involvement often increases productivity are not central here because they are based on faulty analyses in at least two respects: (*a*) ego involvement is generally defined in terms of interest in the task; task-orienting instructions seem designed to decrease interest. In such cases, it cannot be assumed that the motivation of the "ego-involved" subjects is *egocentric*. (*b*) Even when the instructions do not contain this defect, the tasks themselves are not constructed so that they form inherently organized wholes. When the tasks involved are sensible and have objective internal organization, we should expect ego-oriented individuals, since they are more diffuse with respect to the general requirement of task orientation, to behave more diffusely and therefore to perform more poorly than task-oriented individuals. When the tasks have an arbitrary character, as in classical conditioning, ego-oriented individuals should perform "better." In the latter case, the better performance is actually the more inefficient in relation to the task requirements.

An unpublished experiment by Birch[5] appears to provide an ingenious circumvention of the difficulties outlined above. Three different groups of subjects were asked to solve Wertheimer's 9-dot problem. The actual experimenter was one of Birch's students. One group (A) was told that the test measures capacity for graduate work, and that departmental recommendations would be based on the results. In another group (B) this test orientation was hinted at, but cooperation with the experimenter was emphasized. The third group (C) was told that the experimenter was getting the data for his professor, who was a horrible taskmaster; the entire emphasis was on requesting the subject to do his best for the sake of the unfortunate student experimenter. This last type of instruction should result in maintaining the subjects' interest, but on a cooperative, rather than egocentric or competitive level. The results show no significant differences between the groups in number of solutions achieved. However, systematic differences were found in the number of times subjects drew lines to extend beyond the boundaries set by the dots. The groups ranked in the order CBA from most to least. Since going beyond the boundaries set by the dots is the first prerequisite for a solution, it may be that such a measure is a more accurate indication of the degree of meeting task requirements than actual solution. The latter measure may be confounded with other variables, e.g., ability in spatial perception, specific experience, etc.[6]

---

[4] In this connection, see Asch's analysis of Watson and Raynor's experiment on conditioned fear in infants (2, pp. 99–101).

[5] Birch, H. Personal communication, 1953.

Clinical observation also indicates the relevance of evaluation of a patient's motivational structure with respect to the assumed environmental requirements of cooperation and task-oriented interest in his surroundings. Perhaps the most common characteristic of schizophrenics is withdrawal from social relationships. Thus, schizophrenia represents a high degree of psychopathology. Schizophrenics are oriented away from all social requirements. Many psychoneurotics are characterized by hostile competitiveness. Therefore, this is a milder degree of psychopathology, since social relationships are at least considered, but they are not dealt with efficiently.

A rather incisive analysis of the effects of egocentricity on the perception of a social situation was made by Wertheimer (23, ch. 4). In the same place, Wertheimer also tells of the positive effects of changing a competitive relationship into a cooperative one in the case of a game of badminton. Wertheimer's materials suggest many experimental hypotheses in this field, but controlled studies are disappointingly scarce.

To summarize this section, necessarily sketchy because of the scarcity of controlled experimentation, it is proposed that the concept of efficiency can be utilized in the area of motivation. On the assumption that cooperativeness and task involvement are general environmental requirements, it is predicted that inefficient behavior would be a function of the degree of incongruity of an individual's motives with respect to these requirements. It should be clear that congruity and incongruity as descriptive of motivation are parallel to focused and diffuse in descriptions of motor activity. Thus, the concept of efficiency is held constant across all the areas discussed to this point.

EFFICIENCY AS A CENTRAL EGO FUNCTION

Within the framework of general personality theory, efficiency may be regarded as reflecting a central ego function. In the simplest terms, ego is here understood as having to do with the direction of the organism in its interaction with the environment.[7] Efficiency, then, is not a trait in the sense of permanence. Nor is it part of temperament or character, as these terms are commonly understood. Rather, the degree of efficiency of behavior reflects the degree with which the ego manages to focus the organism's attention and action on the task requirements. Efficiency, therefore, should be affected by various environmental changes, as has been indicated above.

---

[6] Of course, both the independent and dependent variables may be somewhat confounded here: the former with anxiety, the latter with the same factors which are thought to disqualify number of solutions as an appropriate measure. The evidence from this experiment must therefore be evaluated with caution.

DOES EFFICIENCY IN MEETING TASK REQUIREMENTS IMPLY EQUATING PSYCHOLOGICAL HEALTH WITH SOCIAL CONFORMITY?

While possible application of the concept of efficiency to everyday situations is beyond the scope of this paper, the question above intrudes itself rather insistently because of the superficial similarity between "meeting task requirements" and conformity to majority opinion or behavior. In fact, however, they have little in common. A task requirement involves an objective description of the structure of a situation, and what is demanded by it, in terms of a total gestalt. Majority opinion and other social forces may be significant aspects of the situation, but are not necessarily all of it. Indeed, efficient behavior may consist in going counter to the majority.

Many examples of this could be culled from everyday life situations. A pertinent experimental illustration is provided by Asch (2, ch. 16) in his work on group pressure. A naive subject is introduced into a group of the experimenter's confederates. The subjects are presented with a group of three lines and with a standard. Their task is to report which one of the three is equal in length to the standard. Under control conditions, this can be done with almost perfect accuracy. In the experimental condition, where the confederates speak first, 20 to 30 per cent of the responses conform to the majority when the latter is instructed to pick the wrong line. How is this behavior to be evaluated? Clearly, it is inefficient with respect to the instructions given by the experimenter. The yielding subjects have allowed other factors to affect their reports. Efficient behavior in this case consists in focusing on the requirement to give the correct response rather than on majority opinion. In relation to psychological health, a subject is healthy in this case to the degree that he resists social pressure.[8] In other situations, of course, overt conformity to majority opinion may be necessary, but these need not be gone into here. In any case, it should be clear that efficiency in meeting task requirements and sheer conformity, or passive adjustment, are not the same thing.

SUMMARY

The main purpose of this paper was to outline a descriptive conception of psychological health and psychopathology, conceived as a continuum. It was proposed that this continuum be defined in terms of the comparative efficiency in meeting task requirements. Analysis indicated that efficiency is a meaningful parameter only when task requirements can be specified objec-

---

[7] For a more complete discussion of the nature and need for a concept of the ego, see (2, ch. 10).

[8] Of course, this may not be true in any individual case. The motivation for remaining independent may be egocentric; contrariwise, yielding may be based on too subtle an evaluation of the situation, rather than on egocentric motives. Nevertheless, one would expect the relationship to hold generally.

tively. When this can be done, it is hypothesized that efficiency (E) is a function of the ratio of focused (F) to diffuse (D) activity, and of productivity (P), or performance: $E = f(F/D, P)$. F activity is defined as activity directly relevant to the task requirements; D activity is that which is irrelevant to the task requirements.

The concept of efficiency was then utilized in an analysis of some of the data on conditioning under anxiety. It was pointed out that the data in this area do not appear to be specific to anxiety. Rather, increased rates of conditioning are found with increasing degrees of psychopathology, or diffusion. Classical conditioning was then analyzed as a situation requiring arbitrary associations, and it was assumed on the basis of existing data that such associations are facilitated in diffuse states. The more rapid conditioning of pathological groups in such conditions was seen as actually inefficient behavior. It was pointed out that the results in intrinsically organized tasks showed normals to be more efficient.

An attempt was also made to analyze some aspects of motivation from the point of view of efficiency. It was hypothesized that egocentric motivation was inefficient because of the general environmental requirements for task orientation and cooperation. Thus, egocentric motivation would lead to poor performance in sensible tasks, as defined above.

Efficiency was conceptualized as a central ego process, subject to change by various situational factors. It was shown that it was not synonymous with conformity. At the same time, no general application to everyday life situations was attempted, except to indicate that the present proposal did not appear to violate current inferences from clinical observation.

REFERENCES

1. ARNOLD, MAGDA B. A study of tension in relation to breakdown. *J. gen. Psychol.*, 1942, 26, 316–346.
2. ASCH, S. E. *Social psychology.* New York: Prentice-Hall, 1952.
3. BITTERMAN, M. E., & HOLTZMAN, W. H. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, 47, 615–623.
4. CAMERON, N., & MAGARET, ANN. *Behavior pathology.* New York: Houghton Mifflin, 1951.
5. CHILD, I. L. Personality. *Annu. Rev. Psychol.*, 1954, 5, 149–170.
6. DUFFY, ELIZABETH. Tensions and emotional factors in reaction. *Genet. Psychol. Monogr.*, 1930, 7, 1–79.
7. FARBER, I. E., & SPENCE, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, 45, 120–125.
8. HALL, K. R. L., & CROOKES, T. G. Studies in learning impairment. II: Psychoneurotic patients. *J. ment. Sci.*, 1952, 98, 273–279.
9. KLEIN, D. B. Abnormal psychology. In H. Helson (Ed.), *Theoretical foundations of psychology.* New York: D. Van Nostrand, 1951. Pp. 698–759.
10. LURIA, A. R. *The nature of human conflicts.* New York: Liveright, 1932.
11. MALMO, R. B., SHAGASS, C., BELANGER, D. J., & SMITH, A. A. Motor control in psychiatric patients under experimental stress. *J. abnorm. soc. Psychol.*, 1951, 46, 539–547.
12. MALMO, R. B., SHAGASS, C., & SMITH, A. A. Responsiveness in chronic schizophrenia. *J. Pers.*, 1951, 19, 359–375.
13. MORGAN, J. J. B. The overcoming of distraction and other resistances. *Arch. Psychol.*, 1916, No. 35.
14. SARASON, S. B., MANDLER, G., & CRAIGHILL, P. G. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, 47, 561–565.
15. SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, 45, 116–119.
16. SPENCE, K. W., & TAYLOR, JANET A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, 42, 183–188.

17. SPENCE, K. W., & TAYLOR, JANET A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1953, **45**, 265–272.

18. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.

19. TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.

20. TELFORD, C. W., & SWENSON, W. J. Changes in muscular tension during learning. *J. exp. Psychol.*, 1942, **30**, 236–246.

21. WELCH, L., & KUBIS, J. Conditioned PGR (psychogalvanic response) in states of pathological anxiety. *J. nerv. ment. Dis.*, 1947, **105**, 372–381.

22. WELCH, L., & KUBIS, J. The effect of anxiety on the conditioning rate and stability of the PGR. *J. Psychol.*, 1947, **23**, 83–91.

23. WERTHEIMER, M. *Productive thinking.* New York: Harper, 1945.

24. WISHNER, J. Neurosis and tension: an exploratory study of the relationship of physiological and Rorschach measures. *J. abnorm. soc. Psychol.*, 1953, **48**, 253–260.

(Received April 5, 1954)

VOL. 62, NO. 3 MAY, 1955

THE PSYCHOLOGICAL REVIEW

STATISTICAL THEORY OF SPONTANEOUS RECOVERY AND REGRESSION

W. K. ESTES[1]

Indiana University

From the viewpoint of one interested in constructing a learning theory, it would be convenient if an organism's habits of responding with respect to any given situation were modifiable only during periods of exposure to the situation. In that case, it would not be unreasonable, prima facie, to hope that all of the empirical laws of learning could be stated in terms of relations between behavioral and environmental variables. Nothing in psychology is much more certain, however, than that orderly changes in response tendencies—e.g., spontaneous recovery, forgetting—do occur during intervals when the organism and the situation are well separated.

How are these "spontaneous" changes to be accounted for? It is easy enough to construct a law expressing some behavioral measure as a function of time, but an unfilled temporal interval never remains permanently satisfying as an explanatory variable. The temporal gap has to be filled with events of some sort, observed or inferred, in the environment or in the organism. The favorite candidate for the intervening position has usually been a postulated state or process, either neural or purely hypothetical, which varies spontaneously during rest intervals in whatever manner is required to account for the behavioral changes. The difficulty with this type of construct is that it is always much easier to postulate than to unpostulate. Few hypothetical entities are so ill-favored that once having secured a foothold they cannot face out each new turn of empirical events with the aid of a few *ad hoc* assumptions.

The approach to time-dependent learning phenomena which will be illustrated in this paper attempts to shift the burden of explanation from hypothesized processes in the organism to statistical properties of environmental events. The very extensiveness of the array of hypothetical constructs—e.g., set, reactive inhibition, memory trace—which now compete for attention in this area suggests that postulates of this type have entered the scene prematurely. Until more parsimonious explanatory variables have been fully explored, it will scarcely be possible either to define clearly the class of problems which require explanation or to evaluate the various special hypotheses that have been proposed.

By "more parsimonious" sources of explanation, I refer to the variables, ordinarily stimulus variables, which are intrinsic to a given type of behavioral situation and thus must be expected to play a role in any interpretive schema.

Vol. 62, No. 2. March 1955- M

---

[1] This paper was prepared during the author's tenure as a faculty research fellow of the Social Science Research Council.

In the present instance we are interested specifically in the way learned response tendencies change during rest intervals following experimental periods. And we note that there are two principal ways in which stimulus variables could lead to modification in response tendencies during rest intervals. The first is the direct effect that changes in the stimulus characteristics of experimental situations from trial to trial or period to period may have upon response probability. The second is the learning that may occur between periods if the stimulating situations obtaining within and between periods have elements in common. The former category can again be subdivided according as the environmental variation is systematic or random.

The random component has been selected as our first subject of investigation for several reasons. One is that it has received little attention heretofore in learning theory. Another is that in other sciences apparently spontaneous changes in observables have frequently turned out to be attributable to random processes at a more molecular level. Perhaps not surprisingly, considerable analysis has been needed in order to ascertain how random environmental fluctuations during intervals of rest following learning periods would be expected to influence response probabilities. It will require the remainder of this paper to summarize the methods and results of this one phase of the over-all investigation.

GENERAL THEORY OF STIMULUS FLUCTUATION

Even prior to a detailed analysis, we can anticipate that whenever environmental fluctuation occurs, the probability of a response at the end of one experimental period will not be the same as the probability at the beginning of the next. If conditioning is carried out during a given period, some of the newly conditioned stimulus elements[2] will be replaced before the next period by elements which have not previously been available for conditioning. Similarly, during the interval following an extinction period, random fluctuation will lead to the replacement of some of the just extinguished stimulus elements by others which were sampled during conditioning but have not been available during extinction. In either case, the result will be a progressive change in response probability as a function of duration of the rest interval.

In order to make these ideas testable, we must state more formally and explicitly the concepts and assumptions involved. Once this is done, we will have in effect a fragmentary theory, or model, which may account for certain apparently spontaneous changes in response tendencies. At a minimum, this formal model will enable us to derive the logical consequences of the concept of random environmental fluctuation so that they may be tested against experimental data. If the correspondence turns out to be good, we may wish to incorporate this model into the conceptual structure of S-R learning theory, viewing it as a limited theory which accounts for a specific class of time-dependent phenomena.

Most of the assumptions we shall require have been discussed elsewhere (8) and need only be restated briefly for our present purposes.

a. Any environmental situation, as constituted at a given time, determines for a given organism a population of stimulus events from which a sample affects the organism's behavior at any instant; in statistical learning theories the population is conceptualized as a set of stimulus elements from which a random sample is drawn on each trial.

---

[2] For reasons of mathematical simplicity and convenience I shall develop these ideas in terms of the concepts of statistical learning theory. It will be apparent, however, that within the Hullian system a similar argument could be worked out in terms of the fluctuation of stimuli along generalization continua.

b. Conditioning and extinction occur only with respect to the elements sampled on a trial.

c. The behaviors available to an organism in a given situation may be categorized into mutually exclusive and exhaustive response classes.

d. At any time, each stimulus element in the population is conditioned to exactly one of these response classes.

On the basis of these assumptions, functions have been derived by various investigators (2, 5, 8, 16, 21) to describe the course of learning predicted for an idealized situation in which the physical environment is perfectly constant and the organism samples the stimulus population on each trial. No idealized situations are available for testing purposes, but the theory seems to give good approximations to empirical learning functions obtained in short experimental periods under well-controlled conditions.

In the present paper we turn our attention from behavioral changes that occur within experimental periods to the changes that occur as a function of the intervals between periods. Correspondingly, we replace the simplifying assumption of a perfectly constant situation with the assumption of a randomly fluctuating situation.[3] Specifically, it will be assumed that the availability of stimulus elements during a given learning period depends upon a large number of independently variable components or aspects of the environmental situation, all of which undergo constant random fluctuation.

Now let us consider the type of experiment in which an organism is run for more than one period in the same apparatus. In dealing with the behavior that occurs during any given experimental period, the total population S_* of stimulus elements available in the situation at any time during the experiment can be partitioned into two portions: the subset S of elements which are available during that period and the subset S' of elements which are not. Under the conditions considered in this paper, the probability of a response at any given time during the period is equal to the proportion of elements in the available set S that are conditioned to that response. Owing to environmental fluctuation, there is some probability j that an element in the available set S will become unavailable, i.e., go into S', during any given interval Δt, and a probability j' that an element in S' will enter S. These ideas are illustrated in Fig. 1 for a hypothetical situation.

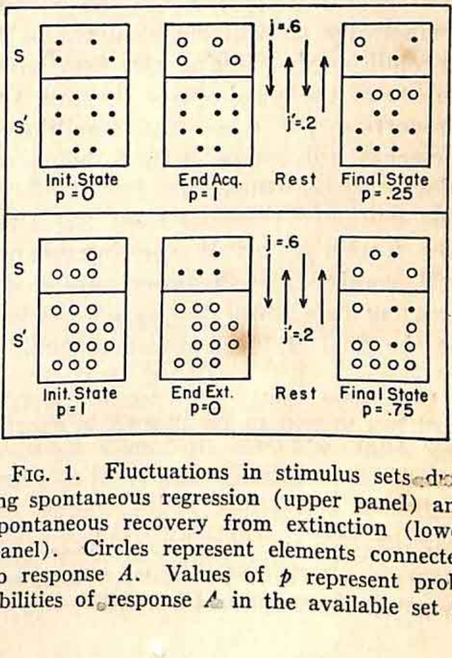


FIG. 1. Fluctuations in stimulus sets during spontaneous regression (upper panel) and spontaneous recovery from extinction (lower panel). Circles represent elements connected to response *A*. Values of *p* represent probabilities of response *A* in the available set *S*.

---

[3] It is possible now to go back and "correct" the functions derived earlier to allow for this random variation, but we will not be able to go into this point in the present paper.

The relevance of the scheme for learning phenomena arises from the fact that both conditioned and unconditioned elements will constantly be fluctuating in and out of the available set S. During an experimental period in which conditioning or extinction occurs, the proportion of conditioned elements in S will increase or decrease relative to the proportion in S'. But during a subsequent rest interval, these proportions will tend toward equality as a result of the fluctuation process.

INTERPRETATION OF SPONTANEOUS RECOVERY AND REGRESSION[4]

The essentials of our treatment of spontaneous recovery and regression will be clear from an inspection of Fig. 1. The upper panel illustrates a case in which, starting from a zero level, conditioning of a given response A is carried out during one period until the probability of A in the available situation represented by the set S is unity. At the end of the conditioning period we will have, neglecting any fluctuation that may have occurred during the period, all of the elements in S conditioned to A and all of the temporarily unavailable elements in S' unconditioned. During the first interval Δt of the ensuing rest interval, the proportion $j = .6$ of the conditioned elements will escape from S, being replaced by the proportion $j' = .2$ of the unconditioned elements from S'. During further intervals the interchange will continue, at a progressively decreasing rate, until the system arrives at the final state of statistical equilibrium in which the densities of conditioned elements in S and S' are equal. The predicted course of spontaneous regression in terms of the proportion of conditioned elements that will be in S at any time following the conditioning period is given by the topmost curve in the upper panel of Fig. 2. The equation of the curve will be derived in a later section.

In an analogous fashion the essentials of the spontaneous recovery process are schematized in the lower panel of Fig. 1. We begin at the left with a situation following maximal conditioning so that all elements are conditioned to response A. During a single period of extinction, all elements in the available set S are conditioned to the class of competing responses $\bar{A}$ and the probability of A goes temporarily to zero. Then during a recovery interval, the random interchange of conditioned and unconditioned elements between S and S' results in a gradual increase in the proportion of conditioned elements in S until the final equilibrium state is reached. The predicted course of spontaneous recovery as a function of time is given by the topmost curve in the lower panel of Fig. 3.

According to this analysis, spontaneous regression and recovery are to be regarded as two aspects of the same process. In each case the form of the process is given by a negatively accelerated curve with the relative rate of change depending solely upon the characteristics of the physical situation embodied in the parameters j and j'. Rates of regression and recovery should, then, vary together whenever the variability of the stimulating situation is modified.

It cannot be assumed, however, that amounts of regression and recovery should be equal and opposite in all experiments. The illustrative example of Fig. 1 meets two special conditions that do not always hold: (a) the condition-

---

[4] The term *spontaneous regression* will be used here to refer to any decrease in response probability which is attributable solely to stimulus fluctuation. It is assumed that over short time intervals, the empirical phenomenon of forgetting may be virtually identified with regression, but that over longer intervals forgetting is influenced to an increasing extent by effects of interpolated learning.

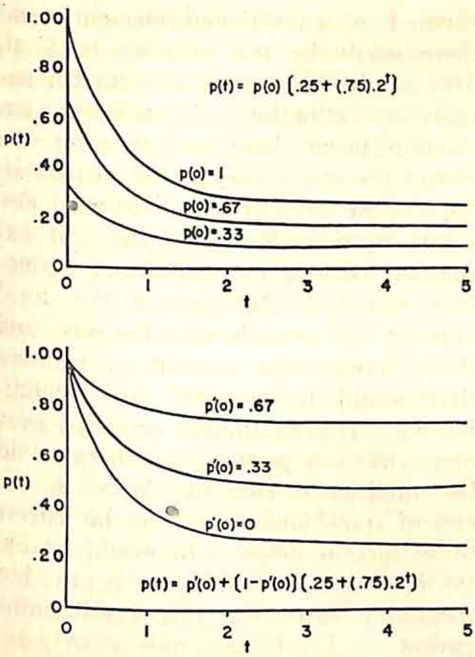


FIG. 2. Families of spontaneous regression curves. In the upper panel the proportion of conditioned elements in S' at the end of conditioning is zero and the proportion in S is the parameter. In the lower panel the proportion of conditioned elements in S at the end of conditioning is unity and the proportion in S' is the parameter.

ing and extinction series start from initial response probabilities of zero and unity, respectively; and (*b*) conditioning and extinction are carried to comparable criteria within the experimental period preceding the rest interval.

PREDICTIONS CONCERNING EFFECTS OF EXPERIMENTAL VARIABLES

Terminal level of conditioning or extinction. If other conditions remain fixed, the level of response probability attained at the end of a single learning period will determine both the initial value and the asymptote of the curve of regression or recovery. For the situation represented by the upper panel of Fig. 1, the curve of conditioning goes to unity, and the predicted course of spontaneous regression is given by the top curve in the upper panel of Fig. 2. If in the same situation, conditioning has been carried only to a probability level of, say, .67, then the total number of conditioned elements will be smaller and the curve of regression will not only start at a lower value, but will run to a lower asymptote, and so on. Similarly, if in the situation represented by the lower panel of Fig. 1, response probability goes to zero during the extinction period, the predicted course of spontaneous recovery is given by the lowest curve in the upper panel of Fig. 3; if extinction terminates at higher probability levels, we obtain the successively higher recovery curves shown in the figure.

Number of preceding learning periods. Increasing the number of preceding acquisition periods would tend to increase the total number of condi-

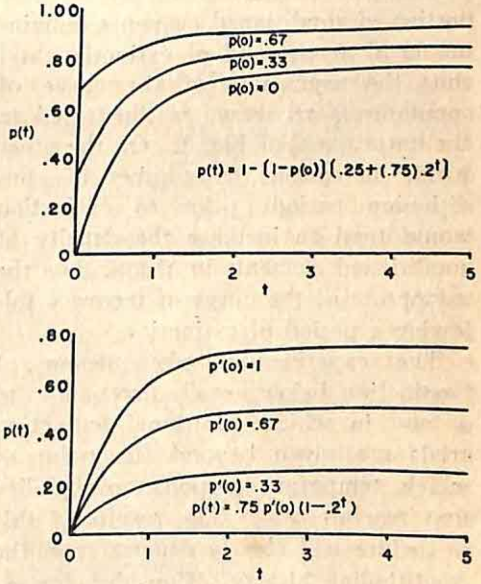


FIG. 3. Families of spontaneous recovery curves. In the upper panel the proportion of conditioned elements in S' at the end of extinction is unity and the proportion of conditioned elements in S at the end of extinction is the parameter. In the lower panel, the proportion of conditioned elements in S' at the end of extinction is the parameter and the proportion in S is zero.

tioned elements in S_* and therefore the asymptote of the curve of regression. If level of response probability at the end of the last acquisition period is fixed at some one value, say unity, then variation in the proportion of conditioned elements in S' yields the family of regression curves illustrated in the lower panel of Fig. 2, all curves starting at the same point but diverging to different asymptotes. This curve family will be recognized as corresponding to the well-known relationship between retention and amount of overlearning, where overlearning is defined in terms of additional training beyond the point at which response probability in the temporarily available situation reaches unity.

Analogous considerations apply in the case of spontaneous recovery. Increasing the number of preceding extinction periods would tend to decrease the proportion of conditioned elements remaining in S' at the end of extinction and thus the asymptote of the curve of spontaneous recovery, as illustrated in the lower panel of Fig. 3. On the other hand, increasing the number of conditioning periods prior to extinction would tend to increase the density of conditioned elements in S' and thus the asymptote of the curve of recovery following a period of extinction.

The experimental phenomenon of "extinction below zero" corresponds to a case in which additional extinction trials are given beyond the point at which temporary response probability first reaches zero. The results of this procedure will clearly depend upon the conditioning history. Consider, for example, the situation illustrated in the top row of Fig. 1. If extinction were begun immediately following the conditioning period, then we would expect extinction below zero to have little effect, for at the end of the first extinction period the set S would be exhausted of conditioned elements and there would be few or none in S' to fluctuate back into S during further periods of extinction. If, however, extinction began long enough after the end of the acquisition period so that an appreciable number of conditioned elements were in S' during the first extinction period, the additional extinction would further reduce the total number of conditioned elements and thus increase the amount of training that would be required for reconditioning. If conditioning extended over more than one period, then there would be conditioned elements in S' at the end of conditioning, and similar effects of extinction below zero would be expected even if extinction began immediately after the last conditioning period.

Distribution of practice. In general, amount of spontaneous regression should vary inversely with duration of the intertrial interval during conditioning, and spontaneous recovery should vary inversely with duration of the intertrial interval during extinction. In each case, the length of the intertrial interval will determine the extent to which the stimulating situation can change between trials, and thus the proportion of the elements in the stimulus population S_* which will be sampled during a given number of trials. These relationships will be treated in more detail in a forthcoming paper (7).

MATHEMATICAL DEVELOPMENT OF FLUCTUATION THEORY

Stimulus fluctuation model. Let the probability that any given element of a total set S_* is in the available set S at time t be represented by $f(t)$, the probability that an element in S escapes into the unavailable set S' during a time interval Δt by j, and the probability that an element in S' enters S during an interval Δt by j'. Then by elementary probability theory we have for the probability that an element is in S at the end of the $(t+1)$st interval Δt following an experimental period:

$$f(t+1) = [1-f(t)]j' + f(t)(1-j).$$

This difference equation can be solved by standard methods (2, 12) to yield a formula for $f(t)$ in terms of t and the parameters:

$$\begin{aligned} f(t) &= \frac{j'}{j+j'} \\ &\quad - \left[\frac{j'}{j+j'} - f(0)\right](1-j-j')^t \\ &= J - [J - f(0)]a^t \end{aligned} \qquad [1]$$

where $f(0)$ is the initial value of $f(t)$; J represents the fraction $j'/j+j'$; and a represents the quantity $(1-j-j')$. Since a is bounded between -1 and $+1$ by the definition of j and j', the probability that any element[5] is in S will settle down to the constant value J after a sufficiently long interval of time, and the total numbers of elements in S and S' will stabilize at mean values N and N', respectively, which satisfy the relation.

$$N = J(N + N'). \qquad [2]$$

Spontaneous recovery and regression. Curves of spontaneous recovery and regression can now be obtained by appropriate application of Equation 1. Let us designate by $p(t)$ and $p'(t)$ the proportions of conditioned elements, and therefore the response probabilities, in S and S' respectively at time t following an experimental period. The set of conditioned elements in S at time t will come in part from the conditioned elements, $p(0)N$ in number, that were in S at the end of the experimental period, and in part from the conditioned elements, $p'(0)N'$ in number, that were in S'. The probabilities of finding elements from these two sources in S at time t are obtained from Equation 1 by setting $f(0)$ equal to 1 and 0 respectively. With these relations at hand we are ready to write the general expression for spontaneous recovery and regression:

$$\begin{aligned} p(t) &= \frac{1}{N}[p(0)\{J-(J-1)a^t\}N \\ &\qquad + p'(0)J(1-a^t)N'] \\ &= p(0)[J-(J-1)a^t] \\ &\qquad + p'(0)(1-a^t)(1-J), \end{aligned} \qquad [3]$$

the parameters N and N' having been eliminated by means of Equation 2.

The functions illustrated by the curve families of Fig. 2 and 3 are all special cases of Equation 3. In the upper panel of Fig. 2, $p'(0)$ has been set equal to 0; in the lower panel, $p(0)$ has been set equal to 1. In the upper panel of Fig. 3, $p'(0)$ has been set equal to 1; in the lower panel, $p(0)$ has been set equal to 0.

EMPIRICAL RELEVANCE AND ADEQUACY

General considerations. The theoretical developments of the preceding sections present two aspects, one general and one specific, which are by no means on the same footing with regard to testability. It will be necessary to discuss separately the general concept of stimulus fluctuation and the specific mathematical model utilized for purposes of deriving its testable consequences.

---

[5] For simplicity, it has been assumed in this paper that all of the elements in S_* have the same values of j and j'. In dealing with some situations it might be more reasonable to assume that different parameter values are associated with different elements. For example, data obtained by Homme (11) suggest that in the Skinner box a portion of the elements should be regarded as fixed and always available while the remainder fluctuate. Application of an analytic method described elsewhere (8) shows that conclusions in the general case will differ only quantitatively from those given in this paper.

The reason why the fluctuation concept had to be incorporated into a formal theory in order to be tested was, of course, the difficulty of direct observational check. Thus for the present this concept must be treated with the same reserve and even suspicion as any interpretation which appeals to unobservable events. This remoteness from direct observation may, however, represent only a transitory stage in the development of the theory. Relatively direct attacks upon certain aspects of the stimulus element concept are provided by recent experiments (1, 21) in which the sampling of stimulus populations has been modified experimentally and the outcome compared with theoretical expectation. Further, it should be noted that the idea of stimulus fluctuation is well grounded in physical considerations. Surely no one would deny that stimulus fluctuation must occur continuously; the only question is whether fluctuations are large enough under ordinary experimental conditions to yield detectable effects upon behavior. The surmise that they are is not a new one; the idea of fluctuating environmental components has been used in an explanatory sense by a number of investigators in connection with particular problems: e.g., by Pavlov (19) and Skinner (22) in accounting for perturbations in curves of conditioning or extinction, by Guthrie (10) in accounting for the effects of repetition, and recently by Saltz (20) in accounting for disinhibition and reminiscence.

Considered in isolation, the concept of stimulus fluctuation is not even indirectly testable; it must be incorporated into some broader body of theory before empirical consequences can be derived. In the present paper we have found that when this concept is taken in conjunction with other concepts and assumptions common to contemporary statistical learning theories (2, 5, 8, 16), the result of the union is a mathematical model which yields a large number of predictions concerning changes in response probability during rest intervals. Once formulated, this model is readily subject to experimental test. Its adequacy as a descriptive theory of spontaneous recovery and regression can be evaluated quite independently of the merits of the underlying idea of stimulus fluctuation.

Spontaneous recovery. Space does not permit the detailed discussion of experimental studies, and we shall have to limit ourselves to a brief summary of empirical relationships derivable from the theory, together with appropriate references to the experimental literature. To the best of my knowledge, the references cited include all studies which provide quantitative data suitable for comparison with predicted functions.

a. The curve of recovery is exponential in form (3, 9, 17) with the slope independent of the initial value (3).

b. The asymptote of recovery is inversely related to the degree of extinction (3, 11).

c. The asymptote of recovery is directly related to the number of conditioning periods given prior to extinction (11).

d. The asymptote of recovery is directly related to the spacing of preceding conditioning periods (11).

e. Amount of recovery progressively decreases during a series of successive extinction periods (4; 13; 19, p. 61).

It may be noted that items *c* and *d* represent empirical findings growing out of a study conducted expressly to test certain aspects of the theory. Many additional predictions derivable from the theory must remain unevaluated until appropriate experimental evidence becomes available, e.g., the inverse relation between asymptote of recovery and spacing of extinction trials or periods, and the predictions concerning "extinction below zero" mentioned in a previous section.

Spontaneous regression. Predictions concerning functional relationships between spontaneous regression and such experimental variables as trial spacing or degree of learning parallel those given above for spontaneous recovery, but in the case of regression there are fewer data available for purposes of verification. The predicted exponential decrease in amount of regression as a function of number of preceding learning periods has been observed in several studies (**6**, **11**, **13**, **14**). Predictions concerning regression in relation to spacing of learning periods have not been tested in conditioning situations, but they seem to be in agreement with rather widely established empirical relationships between spacing and retention in human learning (**15**, pp. 156–158; **18**, p. 508).

Finally, the question may be raised whether there are no experimental facts that would embarrass the present theory. If a claim of comprehensiveness had been made for the theory, then negative instances would be abundantly available. Under some conditions, for example, recovery or regression fails to appear at all following extinction or conditioning, respectively. Since, however, we are dealing with a theory that is limited to effects of a single independent variable, stimulus fluctuation, instances of that sort are of no special significance. Like any limited theory, this one can be tested only in situations where suitable measures are taken and where the effects of variables not represented in the model are either negligible or else quantitatively predictable. And subject to these qualifications, available evidence seems to be uniformly confirmatory. The danger of continually evading negative evidence by *ad hoc* appeals to other variables cannot be entirely obviated, but it may be progressively reduced if we are successful in bringing other relevant independent variables into the theoretical fold by further applications of the analytical method illustrated here.

SUMMARY

In this paper we have investigated the possibility that certain apparently spontaneous behavioral changes, e.g., recovery from extinction, may be accounted for in terms of random fluctuation in stimulus conditions. Taken in isolation, the concept of random stimulus fluctuation has proved untestable, but when incorporated into a model it has led to quantitative descriptions of a variety of already established empirical relationships concerning spontaneous recovery and regression and to the determination of some new ones. A forthcoming paper in which the same model is applied to the problem of distribution of practice will provide further evaluation of its scope and usefulness in the interpretation of learning phenomena.

REFERENCES

1. BURKE, C. J., ESTES, W. K., & HELLYER, S. Rate of verbal conditioning in relation to stimulus variability. *J. exp. Psychol.*, 1954, **48**, 153–161.
2. BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning*. New York: Wiley, in press.
3. ELLSON, D. G. Quantitative studies of the interaction of simple habits: I. Recovery from specific and generalized effects of extinction. *J. exp. Psychol.*, 1938, **23**, 339–358.
4. ELLSON, D. G. Successive extinctions of a bar-pressing response in rats. *J. gen. Psychol.*, 1940, **23**, 283–288.
5. ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
6. ESTES, W. K. Effects of competing reactions on the conditioning curve for bar pressing. *J. exp. Psychol.*, 1950, **40**, 200–205.
7. ESTES, W. K. Statistical theory of distributional phenomena in learning. *Psychol. Rev.*, in press.

8. ESTES, W. K., & BURKE, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.

9. GRAHAM, C. H., & GAGNÉ, R. M. The acquisition, extinction, and spontaneous recovery of a conditioned operant response. *J. exp. Psychol.*, 1940, **26**, 251–280.

10. GUTHRIE, E. R. *The psychology of learning.* New York: Harper, 1952.

11. HOMME, L. E. Spontaneous recovery from extinction in relation to number of reinforcements, spacing of acquisition, and duration of initial extinction period. Unpublished Ph.D. thesis, Indiana Univer., 1953.

12. JORDAN, C. *Calculus of finite differences.* New York: Chelsea, 1950.

13. LAUER, D. W., & ESTES, W. K. Successive acquisitions and extinctions of a jumping habit in relation to schedule of reinforcements. *J. comp. physiol. Psychol.*, 1955, **48**, 8–13.

14. LAUER, D. W., & ESTES, W. K. Rate of learning successive discrimination reversals in relation to trial spacing. *Amer. Psychologist*, 1953, **8**, 384. (Abstract)

15. MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.

16. MILLER, G. A., & MCGILL, W. J. A statistical description of verbal learning. *Psychometrika*, 1952, **17**, 369–396.

17. MILLER, N. E., & STEVENSON, S. S. Agitated behavior of rats during experimental extinction and a curve of spontaneous recovery. *J. comp. Psychol.*, 1936, **21**, 205–231.

18. OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

19. PAVLOV, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep.) London: Oxford Univer. Press, 1927.

20. SALTZ, E. A single theory for reminiscence, act regression, and other phenomena. *Psychol. Rev.*, 1953, **60**, 159–171.

21. SCHOEFFLER, M. S. Probability of response to compounds of discriminated stimuli. *J. exp. Psychol.*, 1954, **48**, 323–329.

22. SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century-Crofts, 1938.

(Received April 18, 1954)

Psychological Review
Vol. 62, No. 3, 1955

THE CONSTANCY OF THE *I–V*: A CRITIQUE OF INTERVENING VARIABLES

JOHN P. SEWARD

University of California, Los Angeles

In the last 25 years American psychologists have become acutely theory-conscious. Developments in physics and philosophy are partly responsible for this state of mind. Operationists and logical empiricists have taught us how "free-floating" constructs can generate pseudo problems and vicious circles. Turning a critical eye on our own scientific activities we have found such dangers not exaggerated.

Psychologists tend to react in either of two ways. One group, the positivists, barricades itself against error by establishing a line of defense at the level of empirical generalization. Brunswik (**3**), for example, would keep the organism empty and devote his efforts to correlating its behavioral adjustments with changes in the environment. A second group, the systematists, finds the atmosphere stimulating to greater theoretical effort. Logical criticism has merely made them look more closely to their conceptual weapons and prepare for a longer campaign. Tolman (**30**) pointed the way with his floor plan for operational behaviorism; Hull (**12, 15, 16**) put the plan to work in building his own system.

The gist of the Tolman-Hull method is too well known to need detailed description here. Essentially, it involves setting up a number of hypothetical constructs, or intervening variables (*I-V*),[1] between stimulus variables (*S-V*) and response variables (*R-V*). Each *I-V* is especially sensitive to a single external or internal condition; combined, they give the organism's behavior. The mathematical function relating a response to its many determining stimuli can thus be broken down into a number of simpler components. It should be noted that there are no gaps in this hypothetical chain. In principle, there is no quarrel between the positivist Skinner and the Hullians. His objections to theories of learning are aimed explicitly at statements concerning "events taking place . . . at some other level of observation and measured, if at all, in different dimensions" (**26**, p. 193). In this sense Hull and Tolman are no theorists. Their *I-V*'s belong on another level of abstraction, it is true, but they refer to the same level of observation.

The issue between positivist and systematist is pragmatic. Does the *I-V* pay off? It is the purpose of the present article to examine the arguments for this type of theory in the light of recent evidence.

THE CASE FOR *I-V*

Intervening variables appear to have two main functions in behavior theory: heuristic and formal. The first function precedes the second and is most fully revealed in Tolman's writings. A construct such as habit, drive, expectation, or demand may help to suggest new relationships to be sought between *S-V*'s and *R-V*'s. It may do so, however, without appearing in the final equation, *R-V* = F(*S-V*). Hull's (**13**, p. 285) argument for a construct of habit based on the time gap between acquisition and use belongs in this category.

---

[1] The terms are used interchangeably here, although possible distinctions between them may prove useful for other purposes (**8, 22**).

Since the heuristic role, vital as it is, can be played behind the scenes, we are here concerned with the formal function of the *I-V*. What use can the theorist make of it in integrating experimental results? Bergmann and Spence (**2**) and Hull (**13**) answer that it is a device for simplifying the otherwise baffling mathematical relationships underlying empirical curves. Just how is this simplification brought about and to what end?

First let us consider a single *R-V*; e.g., the speed of a rat in a runway. We suppose this measure to depend on a variety of conditions: hours of deprivation, amount of reward, length of runway, number of previous trials, pattern of reinforcements, and others as yet unspecified.[2] Our problem is, given certain values of these variables, to predict the rat's running speed. To solve this problem Tolman (**30**) suggests that we vary one *S-V* at a time, holding all others constant, and measure the variations in speed that result. This gives us a set of mathematical functions each relating speed to some single *S-V*. We take each function to define an *I-V* and assume that it remains constant for different values of the other *S-V*'s. On this assumption, although Tolman hesitates to draw the inference, we can presumably combine the several functions by some simple rule—e.g., addition or multiplication—to derive *R-V*.

If the theorist is permitted such an assumption he can hardly deny the empiricist the same privilege. A valid objection, however, is that he begs the very question he set out to answer: how do the *S-V*'s combine? To test the assumption that they do not interact he must vary them not only one at a time but jointly. Such tests may lead to more complicated functions expressing the interaction of two or more *S-V*'s. The point is that in carrying out these tests the system builder will have made as many determinations as the positivist. True, he may be able to express the results in a series of simpler equations, but he can combine them to predict only the outcomes already discovered. As long as he confines himself to a single *R-V* it looks as if the investigator can get along without any formal help from *I-V*'s. Spence (**27**) recognized this point and suggested that it might explain Skinner's lack of interest in theory construction. Simplification by itself is not enough to justify the extra effort. The question is whether these simplified expressions can be put to use.

It is time to face the crux of the matter. What does it mean to ask if an *I-V* "pays off," if it can be "put to use"? The real question is: does it enable us to formulate theorems from which behavior can be deduced? Can we, with the aid of *I-V*'s, hope to predict *S-R* relationships that have never actually been observed? By analogy with the physical sciences it might seem that a truly deductive theory must describe an order of events other than the one to be explained. If this were true, prediction of molar outcomes, aside from simple generalization, would have to be based on molecular laws. Skinner would be right and the systematist should either study physiology or collect many facts. If it is *not* true, how can the *I-V*, without providing more than a vicarious foothold on the molecular substratum, lead to bona fide prediction?

---

[2] To call "number of trials" or "pattern of reinforcements" a stimulus variable may seem unwarranted. As here used the term is synonymous with the more familiar "independent variable" of the laboratory; it is introduced solely for convenience.

So far we have considered but one R-V. Adding a second R-V changes the picture. As Feigl (6) pointed out, if we have m causes and n effects and take a pair at a time, there are mn possible relations to be discovered. By inserting an I-V and finding its relation to each cause and effect in turn, we change the number of determinations to $m + n$. This change is a reduction in all cases except where m or $n = 1$ or where $m = n = 2$. Let us assume, for example, three S-V's and two R-V's. To establish a simple relation between each S-V and each R-V requires six determinations. The same situation with one I-V is shown in Fig. 1A. The number of simple relations to be determined is reduced from six to five. As more S-V's or R-V's are added, the amount of reduction increases at an accelerating rate.[3]

This would be all very well, the reader may point out, if I-V could be directly observed. Since it cannot, its relations can only be inferred from measurements of some R-V as a function of some S-V. How, then, can any saving be effected? To answer this question refer to Fig. 1E. Let F stand for any function relating an R-V directly to an S-V. Suppose we wish to find $R\text{-}V_1 = F_{12}(S\text{-}V_2)$ without having to measure it. One method would be as follows:

Step 1. Measure $R\text{-}V_1$ for different values of $S\text{-}V_1$ and find

$$R\text{-}V_1 = F_1(S\text{-}V_1). \qquad [1]$$

Step 2. By logical and mathematical procedures derive

$$I\text{-}V = \phi_1(R\text{-}V_1) \qquad [2]$$

and

$$I\text{-}V = f_1(S\text{-}V_1). \qquad [3]$$

Step 3. Now measure $R\text{-}V_2$ for different values of $S\text{-}V_2$ and find

$$R\text{-}V_2 = F_2(S\text{-}V_2). \qquad [4]$$

Step 4. In like manner derive

$$I\text{-}V = \phi_2(R\text{-}V_2) \qquad [5]$$

and

$$I\text{-}V = f_2(S\text{-}V_2). \qquad [6]$$

Step 5. By combining equations 2 and 6 algebraically, or, as an approximation, by entering numerical values and fitting a curve, we arrive at a statement of $R\text{-}V_1 = F_{12}(S\text{-}V_2)$.

If in Step 5 equations 3 and 5 were used instead of 2 and 6, the result would be $R\text{-}V_2 = F_{21}(S\text{-}V_1)$. Note that from two sets of experimental observations we can write four S-R functions. If the method is sound it appears that even a single I-V can give the scientist something of the predictive power he seeks.

But the method's soundness depends on a basic assumption: that I-V remains the same whether derived from one set of S-R variables or another. This principle we may call the *constancy of the* I-V.

What does it mean to say that an I-V is constant? Even if a corresponding physical mechanism exists we have no way to identify it. The principle might seem at first glance to be meaningless. It can be stated, however, in terms of the functions by which I-V is defined; when so stated it becomes open to verification.

I shall state the constancy principle in the form of three propositions. Each one states a necessary condition for predicting an unknown function. In so doing it raises two questions: (*a*) the time gap between acquisition and use belongs in this category.

---

[3] This example deals only with isolated variables. Wherever two or more S-V's converge on I-V, there is a possible interaction. Counting each interaction as a separate measurement, we must add three first-order and one second-order interactions, bringing the total number of determinations to nine. Without I-V the total would be 14. Actually the added labor of measuring interactions will vary with the efficiency of experimental designs.

Since the heuristic role, vital as it is, can be played behind the scenes, we are here concerned with the formal function of the *I-V*. What use can the theorist make of it in integrating experimental results? Bergmann and Spence (**2**) and Hull (**13**) answer that it is a device for simplifying the otherwise baffling mathematical relationships underlying empirical curves. Just how is this simplification brought about and to what end?

First let us consider a single *R-V*; e.g., the speed of a rat in a runway. We suppose this measure to depend on a variety of conditions: hours of deprivation, amount of reward, length of runway, number of previous trials, pattern of reinforcements, and others as yet unspecified.[2] Our problem is, given certain values of these variables, to predict the rat's running speed. To solve this problem Tolman (**30**) suggests that we vary one *S-V* at a time, holding all others constant, and measure the variations in speed that result. This gives us a set of mathematical functions each relating speed to some single *S-V*. We take each function to define an *I-V* and assume that it remains constant for different values of the other *S-V*'s. On this assumption, although Tolman hesitates to draw the inference, we can presumably combine the several functions by some simple rule—e.g., addition or multiplication—to derive *R-V*.

If the theorist is permitted such an assumption he can hardly deny the empiricist the same privilege. A valid objection, however, is that he begs the very question he set out to answer: how do the *S-V*'s combine? To test the assumption that they do not interact he must vary them not only one at a time but jointly. Such tests may lead to more complicated functions expressing the interaction of two or more *S-V*'s. The point is that in carrying out these tests the system builder will have made as many determinations as the positivist. True, he may be able to express the results in a series of simpler equations, but he can combine them to predict only the outcomes already discovered. As long as he confines himself to a single *R-V* it looks as if the investigator can get along without any formal help from *I-V*'s. Spence (**27**) recognized this point and suggested that it might explain Skinner's lack of interest in theory construction. Simplification by itself is not enough to justify the extra effort. The question is whether these simplified expressions can be put to use.

It is time to face the crux of the matter. What does it mean to ask if an *I-V* "pays off," if it can be "put to use"? The real question is: does it enable us to formulate theorems from which behavior can be deduced? Can we, with the aid of *I-V*'s, hope to predict *S-R* relationships that have never actually been observed? By analogy with the physical sciences it might seem that a truly deductive theory must describe an order of events other than the one to be explained. If this were true, prediction of molar outcomes, aside from simple generalization, would have to be based on molecular laws. Skinner would be right and the systematist should either study physiology or collect many facts. If it is *not* true, how can the *I-V*, without providing more than a vicarious foothold on the molecular substratum, lead to bona fide prediction?

---

[2] To call "number of trials" or "pattern of reinforcements" a stimulus variable may seem unwarranted. As here used the term is synonymous with the more familiar "independent variable" of the laboratory; it is introduced solely for convenience.

So far we have considered but one *R-V*. Adding a second *R-V* changes the picture. As Feigl (6) pointed out, if we have m causes and n effects and take a pair at a time, there are mn possible relations to be discovered. By inserting an *I-V* and finding its relation to each cause and effect in turn, we change the number of determinations to $m + n$. This change is a reduction in all cases except where m or $n = 1$ or where $m = n = 2$. Let us assume, for example, three *S-V*'s and two *R-V*'s. To establish a simple relation between each *S-V* and each *R-V* requires six determinations. The same situation with one *I-V* is shown in Fig. 1A. The number of simple relations to be determined is reduced from six to five. As more *S-V*'s or *R-V*'s are added, the amount of reduction increases at an accelerating rate.[3]

This would be all very well, the reader may point out, if *I-V* could be directly observed. Since it cannot, its relations can only be inferred from measurements of some *R-V* as a function of some *S-V*. How, then, can any saving be effected? To answer this question refer to Fig. 1E. Let F stand for any function relating an *R-V* directly to an *S-V*. Suppose we wish to find $R\text{-}V_1 = \mathrm{F}_{12}(S\text{-}V_2)$ without having to measure it. One method would be as follows:

Step 1. Measure $R\text{-}V_1$ for different values of $S\text{-}V_1$ and find

$$R\text{-}V_1 = \mathrm{F}_1(S\text{-}V_1). \qquad [1]$$

Step 2. By logical and mathematical procedures derive

$$I\text{-}V = \phi_1(R\text{-}V_1) \qquad [2]$$

and

$$I\text{-}V = \mathrm{f}_1(S\text{-}V_1). \qquad [3]$$

Step 3. Now measure $R\text{-}V_2$ for different values of $S\text{-}V_2$ and find

$$R\text{-}V_2 = \mathrm{F}_2(S\text{-}V_2). \qquad [4]$$

Step 4. In like manner derive

$$I\text{-}V = \phi_2(R\text{-}V_2) \qquad [5]$$

and

$$I\text{-}V = \mathrm{f}_2(S\text{-}V_2). \qquad [6]$$

Step 5. By combining equations 2 and 6 algebraically, or, as an approximation, by entering numerical values and fitting a curve, we arrive at a statement of $R\text{-}V_1 = \mathrm{F}_{12}(S\text{-}V_2)$.

If in Step 5 equations 3 and 5 were used instead of 2 and 6, the result would be $R\text{-}V_2 = \mathrm{F}_{21}(S\text{-}V_1)$. Note that from two sets of experimental observations we can write four *S-R* functions. If the method is sound it appears that even a single *I-V* can give the scientist something of the predictive power he seeks.

But the method's soundness depends on a basic assumption: that *I-V* remains the same whether derived from one set of *S-R* variables or another. This principle we may call the *constancy of the* I-V.

What does it mean to say that an *I-V* is constant? Even if a corresponding physical mechanism exists we have no way to identify it. The principle might seem at first glance to be meaningless. It can be stated, however, in terms of the functions by which *I-V* is defined; when so stated it becomes open to verification.

I shall state the constancy principle in the form of three propositions. Each one states a necessary condition for predicting an unknown function. In so doing it raises two questions: (*a*)

---

[3] This example deals only with isolated variables. Wherever two or more *S-V*'s converge on *I-V*, there is a possible interaction. Counting each interaction as a separate measurement, we must add three first-order and one second-order interactions, bringing the total number of determinations to nine. Without *I-V* the total would be 14. Actually the added labor of measuring interactions will vary with the efficiency of experimental designs.

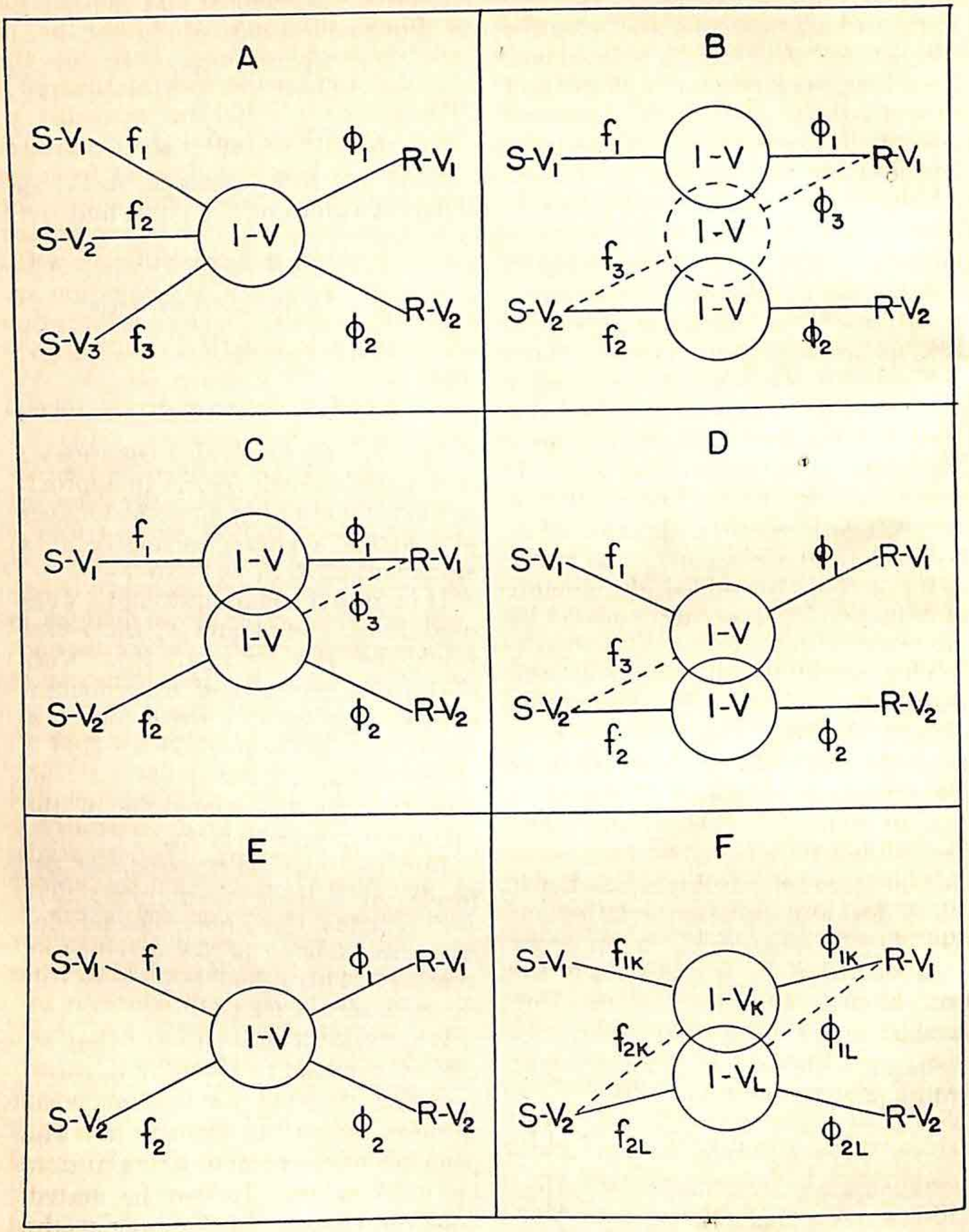


FIG. 1. A. Relations of intervening variable (*I-V*) to stimulus variables (*S-V*) and response variables (*R-V*). Functions relating *I-V* to *S-V* denoted by f; *I-V* to *R-V*, by ϕ. B. Hypothetical situation requiring constancy of *I-V*. Solid lines = known functions; broken lines = functions needed for prediction. C. Same situation with *I-V* constancy as defined by proposition I. D. *I-V* constancy by proposition II. E. *I-V* constancy by both propositions I and II. F. Hypothetical situations *K* and *L* requiring *I-V* constancy by proposition III.

whether the condition is in fact necessary; or (*b*) whether the condition is, in any particular case, met. As to question *a*, I shall be satisfied to answer it by "intuitive" argument rather than rigorous proof, so that we may proceed at once to empirical evidence bearing on question *b*. If I say that a proposition "holds" or is "valid," I shall mean that the answer to question *b* is "yes."

Proposition I: Given a single situation in which the following functions are known: $f_1(S\text{-}V_1) = I\text{-}V = \phi_1(R\text{-}V_1)$ and $f_2(S\text{-}V_2) = I\text{-}V = \phi_2(R\text{-}V_2)$. A necessary condition for predicting $R\text{-}V_1 = F_{12}(S\text{-}V_2)$ is that $I\text{-}V = f(S\text{-}V_2)$ is the same whether derived from $R\text{-}V_2$ or $R\text{-}V_1$.

Figure 1B illustrates the problem. Solid lines represent the situation given; the broken line represents the prediction to be made. Proposition I states that this prediction can be made only if $f_2 = f_3$; i.e., if the situation can be depicted as in Fig. 1C. Obviously if $f_2 \neq f_3$ equation 6 above could not be used in Step 5 to derive F_{12}.

But proposition I, though necessary, is not sufficient for a solution of the problem, since equation 2 is still of uncertain validity. The second requirement may be stated as follows:

Proposition II: Given the problem stated in proposition I, a necessary condition for predicting $R\text{-}V_1 = F_{12}(S\text{-}V_2)$ is that $I\text{-}V = \phi(R\text{-}V_1)$ is the same whether generated by $S\text{-}V_1$ or $S\text{-}V_2$. Starting with Fig. 1B, proposition II requires that $\phi_1 = \phi_3$, a condition which would validate equation 2 for use in Step 5 and would change the diagram to the one shown in Fig. 1D. Finally, if the conditions of both propositions are met, Fig. 1E shows the situation and demonstrates graphically the constancy of $I\text{-}V$ for the two pairs of $S\text{-}R$ variables.

Propositions I and II, taken together and generalized to any number of variables, define $I\text{-}V$ constancy for a single situation. It would be of enormous advantage, both to the builder and to the user of theory, if the principle could be extended to cover situations differing in some respects. A function derived from an experiment on eyelid conditioning, for example, might then be combined with data on conditioned finger withdrawal to provide a new equation applicable to both. Such an assumption is implicit in several of Hull's derivations (e.g., Postulates II, V, VI, XVIA [16]). A more or less formal statement of it follows:

Proposition III: Given two situations, K and L, in which the following functions are known: in K, $f_1(S\text{-}V_1) = I\text{-}V = \phi_1(R\text{-}V_1)$; in L, $f_2(S\text{-}V_2) = I\text{-}V = \phi_2(R\text{-}V_2)$. A necessary condition for making cross predictions in either situation is that the corresponding functions of $I\text{-}V$ are the same in both.

For example, suppose we wish to predict $R\text{-}V_1 = F_{12}(S\text{-}V_2)$. (Propositions I and II are, of course, assumed to hold for K and L separately.) Figure 1F contains the essentials. Proposition III states that the desired function can be derived only if $\phi_{IK} = \phi_{IL}$ and $f_{2K} = f_{2L}$. As in the previous cases this condition is implied by Step 5 of our derivation procedure; it implies in turn that $I\text{-}V_K$ and $I\text{-}V_L$ should be drawn to coincide.

It follows from this discussion that an $I\text{-}V$ defined in terms of several S and R variables can be a powerful tool for making predictions or for deriving new functions. Indeed, the method here outlined may turn out to be an essential part of any theory that makes prediction possible. Its success, however, is clearly contingent on its ability to meet a set of exacting requirements; viz., those laid down in the constancy principle.

HULL'S $_{S}E_{R}$ ON TRIAL

Until recently no psychological theory was far enough advanced to make such tests possible. During the past decade, however, Hull and his students (**7**, **9**, **17**) made a concerted effort in this direction. By applying Thurstone's scaling technique to the bar-pressing latencies (${}_{s}t_{r}$) of rats in successive reinforced trials (N), they were able to define the central construct of his system, reaction potential ($_{S}E_{R}$), as a precise mathematical function of N on the one hand and ${}_{s}t_{r}$ on the other. These equations were then used, much as in the steps outlined above, to derive other postulates. At the same time Hull provided a means of testing the predictive power of his system insofar as it hinges on the constancy of $_{S}E_{R}$.

In this section I shall consider evidence on how well $_{S}E_{R}$ meets the criteria of constancy stated in propositions I to III, and shall take them up in the order III, I, and II.

Evidence on proposition III. Hull's treatment of bar-pressing latencies as a function of number of reinforcements ($\underset{\cdot}{N} = N - 1$) led to the following equation (approximately):

$$_{S}E_{R} = 3.1(1 - 10^{-.03\underset{\cdot}{N}}). \quad [7]$$

Zeaman (**36**) applied the same scaling procedure to starting latencies and running times in a Graham-Gagné apparatus[4] with results as follows:

$$_{S}E_{R}(\text{lat.}) = 3.6(1 - 10^{-.09\underset{\cdot}{N}}), \quad [8]$$

$$_{S}E_{R}(\text{run.}) = 4.7(1 - 10^{-.07\underset{\cdot}{N}}). \quad [9]$$

Seeking to explain the difference in exponents between equations 7 and 8, Zeaman suggested the much larger reward used in his experiment as the most likely factor. But both his own latency data, as he pointed out, and Crespi's (**5**) running time reciprocals as treated by Hull (**15**) showed that amount of reward affected asymptotes rather than exponents. Zeaman's argument implies that the function relating $_{S}E_{R}$ to ${}_{s}t_{r}$ varies with amount of reward, a possibility which, in the light of other evidence to be presented below, should certainly be checked.

Comparing equations 8 and 9, Zeaman attributed the lower asymptote of the latency curve to a longer delay of reinforcement. But this factor should raise the asymptote of equation 7 still higher, since the Yale rats received reinforcement immediately. Rather than attempt to resolve these discrepancies *ad hoc*, we should probably admit that bar-pressing latencies and runway latencies yield *I-V* functions with different constants, then launch a systematic search for the factors responsible.

Greater consistency might well be expected between two sets of bar-pressing data. Yamaguchi (**33**) conducted a study in Hull's laboratory using apparatus practically identical with that of Felsinger *et al.* (**7**). He, too, applied Thurstone's method to his latencies and fitted a curve to the plot of $_{S}E_{R}$ values against successive reinforcements. His equation, to be compared with equation 7, was approximately:

$$_{S}E_{R} = 7.6[1 - (\underset{\cdot}{N} + 1)^{-.2}]. \quad [10]$$

Contrary to expectation, the two equations differ not only in constants but in type. Comparing the experiments we find so many changed conditions that we can only guess which ones are responsible. Differences in maintenance schedule, pretraining procedures, and distribution of trials are equally likely candidates.

Further evidence in the same direction may be found in Hull's Postulate

---

[4] The rats traversed a 3-ft. elevated path from a starting box to a goal box containing food.

XVI, parts A and B. There we find two equations for $_{S}E_{R}$ as a function of unreinforced trials to extinction (n); viz. (approximately),

A. $_{S}E_{R} = 4.0(1 - 10^{-.01n})$, [11]

B. $_{S}E_{R} = .12 \times 10^{.06n} + 2.1$. [12]

Equation 11 was derived indirectly by using a variant of equation 7 in conjunction with bar-pressing data from earlier experiments by Perin (**24**) and Williams (**32**). Equation 12 was derived, also indirectly, by Yamaguchi (**33**), who combined equation 10 with extinction data from his own experiment. It will be noted that equation 11 describes a negatively accelerated, and equation 12 a positively accelerated, curve. Hull's singling out of massed versus "quasi-distributed" reinforcements as the differential condition may or may not be correct. The point that concerns us is that reaction potential as measured turns out to be an exceedingly sensitive instrument; the conditions of its use must therefore be rigidly specified.

Evidence on proposition I. Here we turn from the comparison of two situations, involving unspecified S and R differences, to a comparison of different measures of the same response in the same situation, such as latency, relative frequency, intensity, or resistance to extinction. A difficulty in testing proposition I is to change R-V without also changing the response. Zeaman's study of latencies and running times was not crucial, since he was able to show that the two equations he derived for $_{S}E_{R}$ belonged to two different responses. A recent study by Logan (**21**), however, claims our attention.

Using response times and choice frequencies of two bar-touching responses with different delays of reward, Logan derived three measures of "differential excitatory tendency." His first measure, based on response speed, was a linear function of the difference between latency reciprocals to the short-delay and long-delay bars on single-presentation trials. His second measure, based on "preferential latency" in the same trials, was "the abscissa of the normal probability integral . . . at which the ordinate equals the proportion of times the response speed to the short delay bar is greater than that to the long delay bar" (**21**, p. 302) multiplied by $\sigma\sqrt{2}$.[5] His third measure was derived in the same way from the percentage of choices of the short-delay bar on joint-presentation trials. Instead of plotting all three measures against number of reinforcements and comparing the curves of best fit, Logan tested the inference that the relations among them should be linear.

Logan's analysis would have been more decisive if he had scaled his first measure by Hull's adaptation of Thurstone's method instead of averaging latency reciprocals. The fact that the two estimates based on response speeds and preferential latencies gave a straight-line relationship implies that the former was indeed equivalent to $_{S}E_{R}$, but the matter needs checking to avoid circularity. His second plot, of response speed against percentage of choice measures, also showed linear trends. Departing from theory, however, these trends had steeper slopes than the first one and differed from one part of the experiment to another. Logan also showed that "although the animals developed a 100 per cent choice of the short delay bar, they were responding more rapidly to this bar on only about 90 per cent of the compared latencies" (**21**, p. 306).

Again the question arises, do these

---

[5] σ is roughly equivalent to the standard error of oscillation of $_{S}E_{R}$ defined by Hull in Postulate XII (**16**).

measures refer to the same response? Measures 1 and 2 obviously do, since they represent different ways of treating the same numbers. But is a response to one of two bars presented together the same—i.e., determined by the same external and internal conditions—as a response to the same bar presented alone? The answer must await the verdict of a growing body of experiments comparing discrimination between stimuli simultaneously and successively presented (**1**, **10**, **20**, **25**).

Logan himself appealed to stimulus patterning, itself the product of learning, to account for the divergence of his free-choice and forced-response measures. Another factor may be partly responsible. Theoretically $_SE_R$ to each bar fluctuates from moment to moment in a roughly normal distribution with a constant standard error of oscillation (σ_O). On the assumption that the two distributions of these momentary values ($_SE_R$) are independent, their differences are distributed with $\sigma_D = \sigma_O\sqrt{2}$. The percentage of times the response to one bar is faster than to the other is determined by the ratio of the mean difference to its standard error, which can thus be read from a table of the normal probability integral.

Each single-bar response may be held to represent one $_SE_R$ value, but a choice between two bars may represent more than one $_SE_R$ to each. A rat frequently hesitates, looking from one to the other before he chooses. Such "vicarious trial and error (VTE)" (**23**) samples a number of $_SE_R$'s. It would be plausible, therefore, to say that a single trial provides a difference between the means of two samples and that percentage of choice is determined by the distribution of these differences between means. Letting n stand for the mean size of a sample, we may approximate the standard error of this distribution by $\sigma_{\overline{D}} = \sigma_D/\sqrt{n}$. If, for example, we assume a mean of three VTE's per trial, we should use $\sigma_{\overline{D}} = \sigma_O\sqrt{2/3}$ instead of $\sigma_D = \sigma_O\sqrt{2}$ as on forced trials. It follows that a given differential $_SE_R$ would theoretically produce a larger percentage of choices of the short-delay bar than of faster responses to it, which is what Logan found.

This hypothesis can be checked. In an experiment comparing successive with simultaneous discrimination any condition that increases VTE on the joint-presentation trials should increase the discrepancy between preferential latency and percentage of choice. Tolman (**31**) has suggested a number of conditions that might be varied for this purpose, such as enforced delay at the choice point, difficulty of the discrimination, and the use of normal as compared with brain-injured rats.

As to the constancy of $_SE_R$ in the sense of proposition I, Logan's study confirms it, or at least offers no evidence against it.

Evidence on proposition II. A difficulty in making a direct test of proposition II is that it demands the use of Thurstone's method in connection with two *S-V*'s, whereas so far no study has applied it to more than one. To obtain an expression relating $_SE_R$ to $S\text{-}V_2$ we must, therefore, proceed indirectly, using a second *R-V* and tacitly assuming propositon I. This means that we are not testing proposition II alone but I or II or both. Two sets of data are available for making such indirect tests, one provided by Yamaguchi (**33**, **34**), the other by Perin (**24**). A third study, by Kimble (**19**), gave latencies under conditions comparable with Perin's; they will therefore be used to supplement Perin's data.

Figure 2 is presented to illustrate the test. Here the general symbols

TABLE 1

EMPIRICAL AND THEORETICAL FUNCTIONS TAKEN FROM YAMAGUCHI (34) TO TEST PROPOSITION II

Steps	Empirical	Theoretical
1	(a) ${}_st_r = \mathrm{F}_1(N)$	(b) ${}_SE_R = \phi_1({}_st_r)$
		(c) ${}_SE_R = \mathrm{f}_1(N)$
2	(d) $n = \mathrm{F}_{21}(N)$	(e) ${}_SE_R = \phi_2(n)$
3	(f) $n = \mathrm{F}_2(h)$	(g) ${}_SE_R = \mathrm{f}_2(h)$

Note.—For explanation of symbols see Fig. 1 and 2.

of Fig. 1E are replaced by specific ones standing for the variables used in the two experiments. The hypothesis tested is that ϕ_1 is constant whether derived by varying N or h, the number of hours of food deprivation.

Yamaguchi's purpose was to find an equation for drive (D) as a function of h. His method, similar to the one outlined above, consisted of a series of steps the essentials of which are summarized in Table 1. Step 1 was to measure bar-pressing latencies in 88 reinforced trials, derive Thurstone ${}_SE_R$'s, and fit his (Yamaguchi's) equations b and c, relating them to ${}_st_r$ and N. Equation c appeared as equation 10 above. Step 2 was to extinguish the response in several groups after different numbers of reinforced trials and, by substituting ${}_SE_R$ values from equation c, fit equation e for ${}_SE_R$ as a function of n (see equation 12 above). Note that at this point the soundness of proposition I was implied. In Step 3 the response was trained and extinguished after different amounts of food deprivation. These data, combined with equation e, led to equation g for ${}_SE_R$ as a function of h. By assuming equivalence of ${}_SE_R$ and D at the limit of practice, Yamaguchi finally emerged with the equation for $D = \mathrm{f}(h)$ that appears in Hull's system as Postulate V.

It should now be possible to combine equations b and g and predict the values of ${}_st_r$ to be expected after differing amounts of h. If proposition II holds, and if proposition I can be assumed, the predicted latencies should agree with those actually obtained. Fortunately Yamaguchi published median latencies for trials 72 and 88 of training in order to show that ${}_SH_R$ had reached its asymptote. These data, averaged for the two trials, may be compared with the theoretical latencies in Table 2. Whereas the calculated values, based on the variation of ${}_st_r$ (and n) with N, drop rapidly, especially between 12 and 48 hours, the empirical measures show no trend at all.

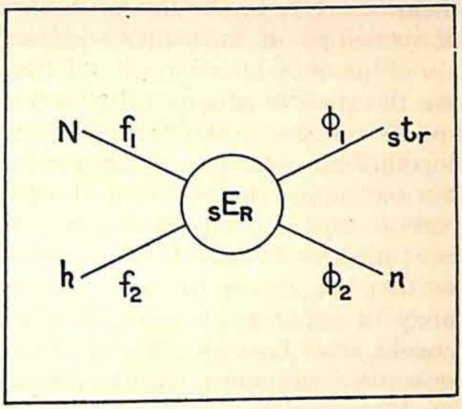


FIG. 2. Experimental situation for testing proposition II provided by Yamaguchi (**34**) and Perin (**24**). N = number of reinforcements; h = hours of food deprivation; ${}_SE_R$ = reaction potential; ${}_st_r$ = latency; n = number of unreinforced trials to extinction.

TABLE 2

MEDIAN BAR–PRESSING LATENCIES (SECONDS) AFTER DIFFERENT DEPRIVATIONS (*h*) AS CALCULATED USING ${}_SE_R$ AND AS OBTAINED BY YAMAGUCHI (**34**)

h	Calculated	Obtained
3	3.56	1.00
12	3.16	1.11
24	0.89	0.93
48	0.16	0.98
72	0.19	1.01

This discrepancy could have been foreseen from Yamaguchi's explanation of his indirect approach. Direct quantification of $_{S}E_{R}$ as a function of h at the limit of learning was "inapplicable here because significant between-group differences in latencies were not obtained except on the earliest training trials" (**34**, p. 114). He attributed this to "the unusual enfeeblement caused by the maintenance procedure." If enfeeblement were solely responsible, however, we should not expect resistance to extinction to vary with hunger as it did. The equalizing of latencies of deprivation groups in the course of training suggests two things: (*a*) As $_{S}E_{R}$ increased it approached a ceiling imposed by physical conditions limiting speed of response. (*b*) If latencies were measured after fewer training trials they might agree more closely with prediction.

With respect to the first implication Cotton (**4**) has reported running times in an 8-ft. runway after 440 reinforced trials, 110 under each of four deprivation periods from 0 to 22 hours. When all test trials were included, clear-cut trends between h values appeared, but when trials with irrelevant responses were excluded, the range of group means for the four drive levels was only about a half second. Although this variation was statistically significant, the author concluded that "the major effect of increased deprivation is to reduce the occurrence of competing responses rather than to increase speed of running per se" (**4**, p. 196). Since bar pressing allows less time for competing responses than running, Yamaguchi's negative finding is understandable.

As to the second implication, the experiments of Perin (**24**) and Kimble (**19**) provide a crude test. Perin gave his rats 16 reinforcements of a bar-pressing response under 23 hours of deprivation, then extinguished four groups at different deprivation periods. Kimble trained his animals, hungry for 24 hours, to push a panel; after 15 reinforced trials he gave tests under different amounts of deprivation in random order. Perin combined his results with the earlier findings of Williams (**32**) in order to state resistance to extinction as a joint function of N and h. By assuming constancy it is possible to write component functions in terms of $_{S}E_{R}$. Since Perin incidentally reported a few latencies and Kimble made them his primary measure, we may again test the constancy assumption.

Our procedure is practically identical with Yamaguchi's and conforms to Table 1 except for the source of the functions. Steps 1 and 2 had already been taken by Hull (**15**, p. 110) in deriving Postulate XVI A, as described above. Equations a, b, and c came from a later experiment (**7**) (see equation 7 above), equation d represented the Perin-Williams data, and equation e emerged as equation 11.[6] Step 3 consisted of using this equation together with Perin's data (equation f) to find $_{S}E_{R}$ values corresponding to his degrees of deprivation (equation g). To do the same for Kimble's data it was necessary first to compute n values for his deprivation periods by equation f (Perin's equation 4 [**24**]). Finally, again invoking proposition III, I used equation b[7] to change both sets of $_{S}E_{R}$'s into theoretical latencies.

---

[6] Propositions I and III seem especially vulnerable here, since equation c was based on one trial a day, while Perin's and Williams' training was massed.

[7] To make equation b consistent with the other equations, I used $_{S}E_{R} = 2.845(_{s}t_{r})^{-.483} - .173$, obtained by adding .426 to Hull's equation 9 (**9**, p. 518) in order to measure from absolute zero.

TABLE 3

MEDIAN LATENCIES (SECONDS) AFTER DIFFERENT DEPRIVATIONS (*h*) AS CALCULATED USING $_sE_R$ AND (AFTER CORRECTION) AS OBTAINED BY PERIN (24) AND KIMBLE (19)

h	Calculated	Obtained*	
		Perin	Kimble
1	6.04	3.82	
2	5.55		3.02
3	5.20	3.62	
8	3.38		2.55
15	1.93		1.55
16	1.76	1.81	
23	1.08	1.00	
24	0.93		1.00

* Perin's scores were divided by 15, Kimble's were multiplied by 5.

The results for all *h* values used in the two experiments are presented in columns 1 and 2 of Table 3.

The actual latencies differed so widely from one study to another that correction factors had to be employed. One of Perin's and one of Kimble's groups, tested after 16 and 15 trials, respectively, gave median latencies of 15 and 0.2 seconds, compared with 0.95 second at the same stage of training in the standardizing experiment (7). I therefore divided Perin's medians by 15 and multiplied Kimble's by 5.[8] The corrected scores appear in columns 3 and 4 of Table 3.

Here the empirical measures, unlike those of Table 1, show a decided slope, although still not so steep as the derived ones. We are not entitled, however, to attribute this trend to the earlier stage of training represented, since both Perin and Kimble trained and tested their animals under conditions shown by Yamaguchi (35) to produce drive stimulus intensity generalization. For this reason, as well as the circuitous nature of the derivation and the large arbitrary correction factors required, our comparison does little more than point up the need of a more adequate test.[9]

It is worth noting that our inquiry into proposition II challenges a basic assumption in Hull's system: that $_sH_R$, as defined in Postulate IV (16), is independent of the strength of drive under which it is acquired. The evidence for this assumption is mainly of two kinds: (*a*) Several investigators (**18, 28, 29**) have found resistance to extinction unrelated to hours of deprivation during training. (*b*) Perin fitted equations to his own and Williams' data on resistance to extinction as a function of number of reinforcements; he found the same exponential constants for training after 3 and 24 hours of deprivation. But Hull quantified $_sH_R$[10] on the basis of reaction latencies with successive reinforcements at a 24-hour drive level. To say that $_sH_R$, so quantified, is independent of drive during training does not follow from the evidence on resistance to extinction.

One reason why $_st_r$ may differ from *n* in this respect was suggested earlier: inertia, or perhaps a lack of competing responses, may keep $_st_r$ from reflecting the growth of $_sH_R$ beyond a certain limit. At high drive levels this ceiling may be reached earlier in training than at lower ones, resulting in a steeper and more sharply bent curve. The more pronounced inflection of equation 10 as compared with equation 7, for example, may well indicate a stronger drive due to the severe maintenance schedule in Yamaguchi's experiment. But what is obviously required is an independent investigation parallel with the original study (7) to determine $_{S}E_{R}$ as a function of N under, say, three hours of deprivation. If the foregoing argument is sound, the resulting function should have a smaller exponential constant than the one in Postulate IV.

---

[8] I estimated Kimble's latencies from his published graph (19).

[9] In developing the functional relationships of $_sE_R$ the Hull group (9) converted Perin's corrected latencies into $_sE_R$'s to plot against *h*. Their corrections involved an error in that they treated Perin's scores as seconds, whereas he actually reported them in minutes. In view of the weaknesses just mentioned, as well as the probable contamination of Perin's data by extinction, a more accurate treatment is not attempted here.

[10] Since $_sE_R$ is treated as a linear function of $_sH_R$, the same quantification applies to both.

It may be objected that the argument, even if sound, is trivial, since the underlying construct may be independent of drive even though its distorted image is not. Such an objection loses sight of the *sine qua non* of a scientific theory, the operational definition of its constructs. If $_{S}H_{R}$ as derived from latencies is found to be dependent on h, the only way its independence can be saved is to redefine it in terms of other operations.

SUSPENDED SENTENCE

To summarize, I have tried to evaluate the usefulness of intervening variables in the explanation, or prediction, of behavior. I argued that the usefulness of an *I-V* depends on the constancy of its relations to *S-V*'s and *R-V*'s as specified in three propositions. I then examined available evidence in order to test the constancy of one *I-V*, Hull's $_{S}E_{R}$, in the only system sufficiently quantitative for such tests to be applied.

What is the outcome? It would obviously be premature to pass judgment on the viability of Hull's construct. Our inquiry merely demonstrates the kind of validation a theory must undergo if it is to make itself useful. This requirement carries several implications for method in theory construction which may be worth stating. They will be couched in terms of animal experimentation for illustrative purposes only.

1. Stability of functional relationships in behavior research demands the utmost uniformity in operations of measurement. While paying lip service to this rule, rat experimenters seldom observe it. As a result, if two findings from different laboratories, or even by different investigators in the same laboratory, confirm each other at all, they can seldom be described by the same equation. A systematic structure can emerge only from rigorously standardized conditions. This means, first, concentration of effort on a few standard situations: enclosed runway, elevated path, Skinner box, single T maze, discrimination box, or jump stand. But more than that, it means agreement on a set of uniform conditions for each situation; e.g., length of runway, work required to press bar, maintenance schedule, pretraining procedures, distribution of trials, age, sex, and strain of subjects.

2. In each situation functional relations are established between several *R-V*'s (e.g., latency, relative frequency, intensity, number of trials to extinction) and several *S-V*'s (e.g., number of reinforcements, hours of deprivation, amount of reward, delay of reward, stimulus intensity).

3. One or more theoretical constructs, or *I-V*'s, are conjectured to mediate these relations. To give such an *I-V* predictive power it must be possible to quantify its relation to a single *S-V* independently on the basis of more than one *R-V*. So far $_{S}E_{R}$ as a function of N has been derived only from latencies and, by way of informal demonstration (**14**), from response frequencies.

4. Similarly, an *I-V* measured by means of a single *R-V* must be independently determined as a function of

more than one *S-V*. So far this has not been done, although it was included in Hull's program (**14**).

If the undertaking is to succeed, certain uniform relations will appear within each situation and between situations. If they do not, new constructs and new scaling devices must be invented and tried.

Advance from simple to complex behavior is implicit in the method outlined. Functions established in runway and Skinner box are tried out with data from T maze and jump stand, from experiments on successive and simultaneous discrimination. Here, too, Hull has led the way with his analysis of simple and compound trial-and-error learning (**16**).

I do not claim this is the way the science of behavior is actually developing. There are those who would deny that it is even an ideal method, holding, with Harlow (**11**), that simple things may be used to explain simpler things but seldom more complex ones. Two issues are likely to be confused here. One, which motivated the present inquiry, lies between positivists and systematists; the other divides rat psychologists and human, or primate, psychologists. Although it happens that the most advanced behavior system to date is closely bound up with experiments on rats, the connection is not necessary. A theorist can conceivably build a psychological system without reference to rat behavior. But whatever his phylogenetic level, if his system is to have any considerable explanatory power he must shape it to meet the kinds of tests here proposed.

REFERENCES

1. BAKER, R. A., & LAWRENCE, D. H. The differential effects of simultaneous and successive stimuli presentation on transposition. *J. comp. physiol. Psychol.*, 1951, **44**, 378–382.
2. BERGMANN, G., & SPENCE, K. W. Operationism and theory in psychology. *Psychol. Rev.*, 1941, **48**, 1–14.
3. BRUNSWIK, E. Organismic achievement and environmental probability. *Psychol. Rev.*, 1943, **50**, 255–272.
4. COTTON, J. W. Running time as a function of amount of food deprivation. *J. exp. Psychol.*, 1953, **46**, 188–198.
5. CRESPI, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.
6. FEIGL, H. Operationism and scientific method. *Psychol. Rev.*, 1945, **52**, 250–259.
7. FELSINGER, J. M., GLADSTONE, A. I., YAMAGUCHI, H. G., & HULL, C. L. Reaction latency ($_st_r$) as a function of the number of reinforcements (N). *J. exp. Psychol.*, 1947, **37**, 214–228.
8. GINSBERG, A. Hypothetical constructs and intervening variables. *Psychol. Rev.*, 1954, **61**, 119–131.
9. GLADSTONE, A. I., YAMAGUCHI, H. G., HULL, C. L., & FELSINGER, J. M. Some functional relationships of reaction potential ($_sE_R$) and related phenomena. *J. exp. Psychol.*, 1947, **37**, 510–526.
10. GRICE, G. R. Visual discrimination learning with simultaneous and successive presentation of stimuli. *J. comp. physiol. Psychol.*, 1949, **42**, 365–373.
11. HARLOW, H. F. Mice, monkeys, men, and motives. *Psychol. Rev.*, 1953, **60**, 23–32.
12. HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
13. HULL, C. L. The problem of intervening variables in molar behavior theory. *Psychol. Rev.*, 1943, **50**, 273–291.
14. HULL, C. L. Research memorandum. (Concerning the empirical determination of the form of certain basic molar behavioral equations and the values of their associated constants.) Unpublished manuscript on file in the libraries of Yale Univer., Univer. of Iowa, Univer. of North Carolina, and Oberlin College, 1946.
15. HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press., 1951.
16. HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
17. HULL, C. L., FELSINGER, J. M., GLADSTONE, A. I., & YAMAGUCHI, H. G. A proposed quantification of habit strength. *Psychol. Rev.*, 1947, **54**, 237–254.

18. KENDLER, H. H. Drive interaction. II. Experimental analysis of the role of drive in learning theory. *J. exp. Psychol.*, 1945, **35**, 188–198.

19. KIMBLE, G. A. Behavior strength as a function of the intensity of the hunger drive. *J. exp. Psychol.*, 1951, **41**, 341–348.

20. LOESS, H. B., & DUNCAN, C. P. Human discrimination learning with simultaneous and successive presentation of stimuli. *J. exp. Psychol.*, 1952, **44**, 215–221.

21. LOGAN, F. A. Three estimates of differential excitatory tendency. *Psychol. Rev.*, 1952, **59**, 300–307.

22. MACCORQUODALE, K., & MEEHL, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, **55**, 95–107.

23. MUENZINGER, K. F. Vicarious trial and error at a point of choice. I. A general survey of its relation to learning efficiency. *J. genet. Psychol.*, 1938, **53**, 75–86.

24. PERIN, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.

25. SALDANKA, E. L., & BITTERMAN, M. E. Relational learning in the rat. *Amer. J. Psychol.*, 1951, **64**, 37–53.

26. SKINNER, B. F. Are theories of learning necessary? *Psychol. Rev.*, 1950, **57**, 193–216.

27. SPENCE, K. W. Mathematical formulations of learning phenomena. *Psychol. Rev.*, 1952, **59**, 152–160.

28. STRASSBURGER, R. C. Resistance to extinction of a conditioned operant as related to drive level at reinforcement. *J. exp. Psychol.*, 1950, **40**, 473–487.

29. TEEL, K. S. Habit strength as a function of motivation during learning. *J. comp. physiol. Psychol.*, 1952, **45**, 188–191.

30. TOLMAN, E. C. Operational behaviorism and current trends in psychology. *Proc. 25th Ann. Celebr. Inaug. Grad. Stud. Univer. Southern California*, 1936, 89–103.

31. TOLMAN, E. C. The determiners of behavior at a choice point. *Psychol. Rev.*, 1938, **45**, 1–41.

32. WILLIAMS, S. B. Resistance to extinction as a function of the number of reinforcements. *J. exp. Psychol.*, 1938, **23**, 506–522.

33. YAMAGUCHI, H. G. Superthreshold reaction potential ($_sE_R$) as a function of experimental extinction (n). *J. exp. Psychol.*, 1951, **41**, 391–400.

34. YAMAGUCHI, H. G. Drive (D) as a function of hours of hunger (h). *J. exp. Psychol.*, 1951, **42**, 108–117.

35. YAMAGUCHI, H. G. Gradients of drive stimulus (S_D) intensity generalization. *J. exp. Psychol.*, 1952, **43**, 298–304.

36. ZEAMAN, D. An application of $_sE_R$ quantification procedure. *Psychol. Rev.*, 1949, **56**, 341–350.

(Received June 1, 1954)

Psychological Review
Vol. 62, No. 3, 1955

PLEASURE–SEEKING BEHAVIOR AND THE DRIVE–REDUCTION HYPOTHESIS[1]

JUDSON S. BROWN

State University of Iowa

During recent years, it has become increasingly apparent that notoriety, if not fame and fortune, can be achieved by repeatedly and vociferously attacking the drive-reduction hypothesis of reinforcement. Just why this hypothesis should have been singled out for special attention is not entirely clear. Perhaps it is indeed strikingly inferior to other conceptions. Perhaps it is one of the few whose formulation is specific enough to make attacks feasible. Or perhaps the onslaughts are manifestations of reaction formation by those whose drives have been reduced through the writing of other polemics. In any event, it must be admitted that these crusades have had a salutary influence. The defenders of the faith have been forced to think more clearly, to avoid circularities of definition, and to perform relevant supporting experiments.

Since these criticisms appear rather frequently in the pages of our psychological journals and books, there is always the danger that such half-truths as they may contain will mislead the casual reader, and that any residual virtues of the drive-reduction assumption will be needlessly obscured or even forgotten. The purpose of this paper, therefore, is to re-examine some of these contentions and to re-emphasize some relatively neglected facts and theoretically plausible mechanisms that bear importantly on the issue. No attempt will be made, however, to defend the notion that drive reduction is, in addition to stimulus-response contiguity, a necessary mechanism for all learning. The concern of this discussion is not with the drive-reduction hypothesis of learning per se. Instead, attention is focussed on the problem of whether behavior commonly described as "pleasure seeking" or "stimulation seeking" constitutes a genuine exception to the drive-reduction interpretation, and whether, without undue strain, such behavior can be coordinated with that conception.

Surveys of the history of the drive-reduction view (22, 28) show that the arguments of its opponents have been not only vigorous, but well diversified. Some writers, placing their trust in the principle of parsimony, have insisted that stimulus-response contiguity alone is sufficient to account for the growth of associative bonds, additional factors of decreased drive or tension being unnecessary. Others have held that the effects attending a consummatory response cannot act backward in time to affect the strength of a connection, or perhaps that "pleasant" consequences alone can produce learning, or even that learning can take place in the absence of both motivation and motivation reduction.

The criticism most frequently raised of late, and the one we shall be concerned with here, has been phrased in a variety of different ways. In all of them, however, a central role has been assigned to the observation that *a response may be learned or maintained even though marked increases in stimulation occur while it is in progress or even after its termination.*

---

[1] Presidential address delivered at the twenty-sixth annual meeting of the Midwestern Psychological Association in Columbus, Ohio, 1954. The author is indebted to Drs. I. E. Farber and Kenneth W. Spence for their constructive criticisms of the manuscript.

In one of the variations on this theme, emphasis is placed on the fact that organisms sometimes exhibit behavior that takes them from an initial state of relative quiescence into one marked by heightened activity and strong stimulation (13). The individual who deserts the quiet comfort of his easy chair for a lurching, noisy ride on a roller coaster is regarded as epitomizing this behavior. He is described as seeking strong stimulation. Since the increased stimulation follows closely upon the response of getting into the roller coaster, and since that increase is sometimes equated with a rise in drive, a problem is said to be raised for the drive-reduction view. In essence, this is the problem of an organism's exhibiting behavior resulting in so-called pleasurable or positive stimulation. It is the problem of why, if the drive-reduction view is correct, men and animals ever behave so as to obtain more stimulation rather than less.

The same basic criticism is sometimes expressed as a dissatisfaction with the explanatory use by drive-reduction theorists of the concept of secondary reinforcement. Hilgard (9), for example, points out that the phenomenon of secondary reinforcement is normally marked by an increase rather than a decrease in stimulation. If stimulus reduction is identified with drive reduction, then a rise in stimulation implies an increase in drive. As a consequence, appeals to secondary reinforcement to explain learning are said to lead to a weakening of the generality of the drive-reduction interpretation.

Still a third formulation of this objection arises from the finding that monkeys, when confined in an opaque-walled box, can learn manual responses for which no reward is given other than that of being allowed to look out of a window of the box (1). According to one interpretation of this result (7), the stimulation provided by the laboratory room beyond the box provides, or arouses, a drive. Consequently, if responses are actually learned when followed by such a drive increase, their acquisition cannot be ascribed to drive reduction. In short, if a response is followed by an increase in stimulation, irrespective of what happens then or later, and regardless of previous experiences with such stimulation, the drive-reduction concept is declared helpless to account for the learning of that response.

A fourth and final variation of this anti-drive-reduction motif rests on the supposed fact that organisms persist for unreasonably long periods in "behavior which is clearly punishing, or, at least more punishing than rewarding" (22, p. 546). The hypothetical author who is described as writing for a lifetime even though the completion of each opus is punished by an editor's rejection slip, and the masochist, are popular examples cited to support this notion. Though the cataloguing of such reactions as "more punishing than rewarding," is obviously unreliable, since hidden rewards may be operating, and "unreasonable persistence" is also difficult to assess, the belief in the existence of such persistent maladaptive behavior is apparently widespread.

To recapitulate, the drive-reduction assumption has been declared deficient for the following reasons: (*a*) Organisms seek out and approach situations or perform actions involving an increase rather than a decrease in stimulation. (*b*) The drive-reduction theorist is inconsistent if he tries to explain the learning of some responses by invoking the principle of secondary reinforcement with its attendant stimulation increments. (*c*) On the assumption that visual stimulation is drive arousing in monkeys, the acquisition of responses "rewarded" by such stimulation is evidence against the drive-reduction concept. (*d*) Organisms are supposed to persist unreasonably in behavior regarded by observers as more punishing than rewarding. In all of these criticisms, stress is laid on the fact that *when increases in stimulation occur after a response has been initiated or after it ceases, the tendency to perform that response is not always weakened, and may even become stronger.* This failure of increased stimulation to function as a kind of "negative reinforcer" is proposed as a major stumbling block for the drive-reduction interpretation.

As an initial step in attempting to answer these criticisms, it may be noted that the proponents of the above views, in order to stimulate thought and present the best possible case for their position, have often emphasized rather bizarre and extreme examples. But the problem raised by these examples is not confined to the bizarre and the abnormal. The phenomena are typical of the simplest and most conventional kinds of behavior. One need not turn to the masochist for pertinent data; they abound in nearly all reactions. Though this point seems to have escaped notice, *almost every overt response, and many a covert one, is accompanied by an increase in stimulation.* In most instances, the level of stimulation provided by proprioceptive feedback and by fluctuating external cues must be higher during the response than either before or after. When a rat is first placed in a maze, it may not move at all for a while. When it does move, however, its reactions are likely to provide an increase in the level of both internal and external stimulation. But if stimulation is, in general, more intense at a later phase of the response than at the moment of its initiation, a problem arises as to whether this constitutes a rise in drive and hence an embarrassment for the drive reductionist. Clearly, this is the same problem that lies at the core of the criticisms that have been mentioned, and just as clearly, it too requires a carefully considered analysis.

Granting, then, that a great many responses, though probably not all, are attended by heightened stimulation, by what stratagems can these stimulus increments be coordinated with the suppositions of the drive-reduction view? One alternative might be to assume that *the sensory consequences of most responses are practically never intense enough to provide increments to the drive level.* But this assumption seems unsatisfactory for the following reasons: (*a*) It implies that when stimuli really are strong, a drive results, an implication that reintroduces the problem to be solved. (*b*) It tacitly assumes that when stimuli do produce a drive the response should be abandoned or at least not learned. (*c*) It raises the further problem of how to specify the precise level of intensity at which stimuli become or induce drives.

A second avenue of escape from this dilemma has been opened by one of Miller and Dollard's proposals (15). They suggest that *heightened stimulation following a response is not incompatible with the drive-reduction position because such stimulation may lower the total stimulation to a level below that prevailing initially.* To use their example, an individual lost in a forest at night does get more stimulation when he sees a light in a cabin window. But seeing the light may actually decrease the totality of stimulation by eliminating internal excitations arising from anxiety or apprehension. Although this solution has substantial merit, the range of cases to which it applies appears to be relatively limited. Moreover, since the rise in stimulation is closer to the response than the decline, a problem remains as to why the decline should override the "punishing" effect of the more immediate rise.

A third possibility, the one to be defended here, is simply that *an increase* *in stimulation, even when relatively intense, need not always be categorized as an increase in drive.* That is, it is sometimes profitable to treat strong stimuli as drives in our formulations for predicting behavior, and sometimes not. The adoption of this view makes necessary the abandonment of the position that each and every stimulus necessarily has drive properties simply because it is physically intense. It means that the definition of a drive must include more than measurements of, or inferences about, the sheer physical strength of a stimulus. It means that a given stimulus may have drive value for some individuals, but not for others; or for one individual at one time, but not at another. This is apparently essentially the same point that Skinner (23) makes when he asserts that a drive is not a stimulus.

With respect to the problem under consideration, this suggestion, that strong stimuli need not always engender drive, has numerous implications. Generally speaking, it makes imperative a careful re-examination of all instances in which strong stimulation is supposed to present a paradox for the drive-reduction view. Specifically, it leads to the conclusion that whenever the jostling and noise of the roller coaster can be disqualified as drive increments, the occurrence of the behavior does not constitute an anomalous example. The masochist who beats himself with a hammer may not be providing sensory excitation whose role as a behavior determinant is covariant with drive. In all these instances, the increase in physical stimulation is present, but the presumed increase in drive, along with its paradoxical implications, vanishes.

Merely to assert, however, that strong stimuli do not always function as drives does not contribute much to the resolution of our difficulties. We need to know when, and under what conditions, stimuli will or will not have drive-like effects. We need to list specific factors other than stimulus intensity which must be taken into account if a maximally useful definition of stimulus-induced drive is to be achieved. Whether the resulting definition is, in fact, useful, will depend on whether certain agreed-upon behavioral consequences are found to accompany or to follow the presentation or removal of the stimulus. Consequences such as the intensification of behavior or the occurrence of learning might serve adequately in this capacity, though no firm decision need be made here.

In attempting to specify these additional factors, let us first consider those which lead to the conclusion that a given stimulus, though strong, will *not* operate as a drive.

One factor capable of altering the drive properties of a stimulus is the amount and kind of previous experience that the organism has had with that stimulus. Thus *a physically intense stimulus tends not to produce a drive if an organism has been exposed on successive occasions to gradually increasing values of that stimulus and if each presentation has been followed by some kind of drive reduction.* Some of the most convincing evidence to support this contention comes from the conditioning studies of Pavlov. He found that stimuli such as strong electric shocks, wounding of the skin severe enough to draw blood, and cauterization of the skin lost their noxious aspects if, when used as conditioned stimuli for an alimentary response, their intensity was gradually increased. The process of neutralization was apparently quite complete, since according to Pavlov's report ". . . not even the tiniest and most subtle objective phenomenon usually exhibited by animals under the influence of strong injurious stimuli can be observed in these dogs. No appreciable changes in the pulse or in the respiration occur in these animals . . ."

(20, p. 30). Incidentally, the fact that the dogs in these experiments were fed for a few seconds after each stimulation appears to be especially significant. Prior to the conditioning training, each of the stimuli apparently had drive value. Marked alterations in respiration and pulse, as well as violent motor reactions directed toward removal of the stimuli or escape from them, were observed. That a complete tolerance was developed for these strong stimuli may be taken as evidence for their loss of drive-arousing capacity. According to Liddell (11), the major findings in these studies have been amply corroborated in studies at the Cornell Behavior Farm.

Further striking evidence to support the view that stimuli normally describable as painful or drive arousing can lose these attributes and become neutral or even "pleasant" if presented at progressively increasing strengths and followed by primary rewards is to be found in Masserman's studies of "experimental masochism." In these experiments, cats were trained to press a switch to obtain food. Air blasts of gradually increasing intensity were then administered each time the switch was depressed. Under these conditions, the air blast can be raised to an intensity that ". . . would be tolerated by no 'normal' cat and yet, because it is now part of the signal-complex leading to feeding, the experimental animal not only is not disturbed by it but, even in the absence of the food reward itself, frequently works the switch to experience the air blast as a subsidiary goal" (12, p. 57). That results of this kind are not restricted to animal subjects is apparent from Slutskaya's study (24) in which infants who were pricked with a needle and then fed came to exhibit anticipatory eating responses to the sight of the needle.

In addition to these experimental data, there are abundant anecdotal illustrations, which, though less well documented, also tend to support this contention. Straight whiskey, when first ingested, typically elicits rather violent defense reactions. Because of this, the novice drinker usually begins with sweet liqueurs, "pink ladies," and wines, and slowly works his way through a series of beverages characterized by the gradual disappearance of cola and ginger-ale additives. Finally, only plain water or even nothing need be mixed with the raw product. To the hardened drinker, straight whiskey does *not* taste bad—not bad at all! (It is thus that a product euphemistically labeled "neutral spirits" becomes indeed psychologically neutral.) That the ingestion of alcoholic beverages is often followed by the reduction of one or more drives is common knowledge. It seems needless, therefore, to speculate here as to which drive, when reduced, provides the principal support for this process of neutralization.

Closely parallel steps seem to be involved in acquiring a taste for highly peppered sauces and in learning to eat the kinds of cheeses that Henning would certainly have located at the putrid corner of his olfactory prism. In each instance samples of the commodity, diluted to the point of toleration, are presented initially and are followed by gradually strengthened versions until a terminal level, limited apparently only by the manufacturer's ingenuity, has been reached.

This process of "demotivating" a strong stimulus seems to have much in common with the phenomenon of *negative adaptation,* especially as it has been described by Guthrie (5). Though he would doubtless not regard the drive-reduction aspect as necessary, he might agree that the task of adapting the horse to the feel of saddle and rider by gradually increasing the weight on its back would be facilitated by the frequent administration of lumps of sugar and crisp Washington apples.

Under some conditions, it might be predicted that the intense proprioceptive stimulation and muscular strain due to prolonged work should have drive-like effects. But if an organism gets appropriately reinforced training, it can acquire a tolerance for the stimulative effects of repetitive muscular effort that is little short of astounding. Rats and pigeons can be trained to make hundreds of responses for a single bite of food if the percentage of reinforcement is high initially, and if the reduction in frequency of reinforcement with further trials is sufficiently gradual. In such instances, apparently, the stimulation accruing from a multitude of successive reactions does not function as a drive, since behavior such as resting, though followed by the cessation of such stimulation, is not strengthened.

Having noted that, in certain learning situations, strong stimuli tend to lose their drive-arousing functions, we may now inquire as to what mechanisms may be involved in this process.

One such mechanism appears to be the transformation of a drive-arousing stimulus into a cue for specific responses appropriate to the situation. In effect, *a stimulus loses its power to act as a drive as it acquires strong associative connections to well-integrated, directed bits of behavior.* For the albino rat, a bright light has demonstrable drive properties. Responses instrumental in escaping from the light, or in turning it off, are readily acquired. But if the light is located near food, and if, through appropriate training, it becomes associated with reactions of approaching and eating, its original drive characteristics will be diminished.

This general notion, that the drive function of a powerful stimulus declines as it acquires new, behavior-directing properties is supported by several lines of evidence. Hebb, for example, makes essentially this point when he asserts that ". . . potentially painful stimulation is not disruptive (that is, not pain) when it forms part of a stimulation that has well-organized central effects" (8, p. 199). In studies of widely distributed periodic reinforcement, stimuli arising from muscular tension and fatigue tend, early in learning, to prevent further responding. But after learning is well advanced, they function as cues for additional reactions. The drive aspects of the strong olfactory and gustatory stimuli arising from whiskey disappear as those cues become associated closely with elbow-bending and swallowing responses. The increase in pleasant affect found by Peters (21) to accompany the making of positive responses to initially unpleasant stimuli may be a comparable phenomenon, and there are a number of parallels between his judgmental theory of feeling and the position taken here.

An unpublished experiment by Miller and Davis[2] illustrates nicely both the manner in which a strong stimulus can acquire associative properties and the importance of making the increases in its intensity gradual if the process of demotivation is to be successful. In their study, rats were first trained to traverse a 2-meter alley for food. Momentary shocks were then administered at a distinctively marked section in the middle of the alley. When shock intensity was strengthened gradually on successive trials, the animals continued to dash rapidly down to the end containing food. But when shock was increased suddenly, the phenomenon of continued running did not appear. Seemingly, when the shock was increased in small steps, it acquired a tendency to elicit running rather than stopping. By becoming associated with the approaching response it relinquished its power to act as a drive. On the basis of this evidence, then, it would appear that drive-arousing stimuli tend to become neutralized as they become converted to cues having strong associations with well-integrated, ongoing responses.

---

[2] Miller, N. E., & Davis, M. The influence of the positions of reward and punishment in the response sequence. Personal communication, 1953.

A second mechanism of importance in effecting the neutralization of strong stimuli during learning involves the acquisition by such stimuli of secondary reinforcing power. The presumption here is *that stimuli lose their driving potentialities, in part at least, as they acquire the capacity to reinforce antedating responses.* The loud buzz produced by the food-delivery magazine in a Skinner box may at first be drive arousing and disturbing to the naive rat. But after being repeatedly paired with food it gains the capacity to reinforce completely new acts and its manifest driving aspect is reduced.

Keller and Schoenfeld (10) stress this mechanism in their analysis of human aversions. As has been done here, they note that originally unpleasant stimuli may lose this character and become acceptable, or even sought after, as a consequence of training. This modification of the properties of the stimulus they attribute not to the outweighing of the aversion by a conflicting drive, but to the aversive stimulus' becoming a secondary reinforcement. They apply this interpretation to masochism, as well as to laboratory studies in which electric shocks administered during reinforced bar-pressing trials acquire the capacity to produce a rapid acceleration in response rate during extinction.

Precisely what happens when a stimulus becomes transformed from a drive arouser to a secondary reinforcer is not clear. One interpretation, which is consistent with the drive-reduction view, would stress the kind of response that becomes associated with the stimulus when it acquires its reinforcing power. If that response is capable of interfering with, or reducing the drive condition of the moment, the net effect of the secondary cue will be an actual diminution in drive. Insofar as this holds true, it means that any inherent drive-arousing power of a secondarily reinforcing stimulus diminishes as it achieves the strength to elicit a drive-reducing reaction. An interpretation of this sort has been proposed for the special case of acquired drives by Miller and Dollard (15). For them, cues can act as learned rewards or secondary reinforcers when they acquire a tendency to inhibit or relax a response that is functioning as a drive. This possibility, that some kind of drive-reduction process underlies the phenomenon of secondary reinforcement, has also been expressed by Mowrer (16), Osgood (18), and Farber (3).

Evidence to support this view comes from diverse sources. Spragg (25), for example, has shown that the intense drive resulting from the withholding of drugs in addicted monkeys can be temporarily reduced by a hypodermic injection of saline solution. The sight and feel of the needle seem to elicit responses of relaxing that are incompatible with whatever states or reactions characterize the condition of abstinence. Wolff and Goodell (29) have found that the administration of a placebo falsely described as aspirin can produce an actual increase in a subject's threshhold for pain. And Gantt (4) has observed that anxiety symptoms in neurotic dogs are completely inhibited by sexual stimulation even prior to orgasm.

It thus appears that learning plays a significant part in the "demotivation" of strong stimuli by leading to the formation of close associations between those stimuli and ongoing activities and by building up the capacity of those stimuli to serve as reinforcers.

In addition to the role played by learning, there are at least two innate processes capable of accomplishing the same result, at least temporarily. One of these is the familiar mechanism of *sensory adaptation.* Instances of this are common in the visual and olfactory modalities, but more striking ones are to be found in research on cutaneous pain. Stone and Jenkins (26), for instance, in a detailed review of such studies, report frequent confirmations of the phenomenon of adaptation to pain. In experiments employing needle algometers, complete pain adaptation was found in approximately 80 to 100 per cent of the instances in which it was studied.

The second innate mechanism may be described, for want of a better label, as *competitive stimulation.* This refers to the fact that the effective strength of pain-arousing stimuli can be weakened by the simultaneous presentation of other stimuli. Hardy and his collaborators (6) state that the pain threshold for one region of the body can be raised as much as 35 per cent by intense stimulation of some other region, a finding confirmed by Parsons and Goetzl (19). Hebb (8) also remarks that stimuli normally capable of provoking pain, and hence presumably drive, may fail to do so when combined with other sources of stimulation. Just how widespread the action of this mechanism may be is uncertain, but it is a common observation that pain can often be reduced by tensing the muscles or by gritting the teeth.

These then are some of the qualifying factors which, when incorporated into a definition of stimulus-induced drive, lead to the decision that certain stimuli, even when physically intense, will not have drive-like effects on behavior. Referring back to the arguments of the anti-drive-reductionists, it should now be apparent that many of the supposed paradoxical effects of increased stimulation disappear in the light of these suggestions. A secondary reinforcer does involve an increase in stimulation, but not necessarily an increase in drive, and it may, in fact, evoke responses which actually reduce drive. The laboratory world, as seen from a window in a box, represents heightened stimulation for a monkey though not inevitably heightened drive. The masochist stimulates himself in a manner that could be painful and drive arousing for others, but not for him.

The solution we have proposed to the problems raised by the anti-drive-reductionists may appear to have achieved its goal at some sacrifice of objectivity. To assert that the drive-arousing potential of a stimulus is not necessarily correlated with its physical intensity is to maintain, of course, that certain characteristics of the organism must be listed among the elements of a definition of stimulus-produced drive. It means that a particular stimulus will not be drive arousing if an organism has had specific kinds of learning experiences with it, if his sensory receptors are in such and such a state of adaptation, or if he is being simultaneously titillated by certain other strong stimuli. This requirement, that the specification of stimulus-aroused drive include additional information about the state of the organism and its history, does to be sure, add complexity to the drive problem, but it presents no insurmountable obstacles. Drive states resulting from operations of deprivation are subject to the same kinds of restrictive specifications. A 24-hour period of food privation means one thing if the subject is old and obese, and something different if the subject is young and slender; it means one thing if the subject has been on an ad libitum diet for weeks, and something else if he has recently undergone a prolonged deprivation schedule. Drive level, defined either in terms of deprivation time or stimulation conditions, can be maximally useful only if additional data about an organism's previous experience, momentary states, and responses are taken into consideration.

In the preceding discussion, some of the conditions under which stimuli would *not* be expected to have drive-like effects have been examined in detail. Before concluding, it seems appropriate to consider briefly some instances in which stimuli *would* be expected to exhibit motivating effects. These are cases in which all the requirements of the definition, as amended by statements about previous experience, sensory adaptation, and so on, have been satisfied. But to grant that intense stimuli can qualify as contributors to drive level is to raise the original problem with which this discussion began. An admission has been made that responses followed by intense stimulation may be learned or perpetuated, at least when the stimuli are deleted from the category of drives. But can it be admitted that responses may be learned or maintained when their onset is followed by stimuli that cannot be stricken from the class of drives?

Either an affirmative or negative answer can be given to this question, depending on the nature and strength of responses evoked by the stimulus and on the relation between the time of drive increase and the time of occurrence of the response. The discussion of this matter can be conveniently organized with respect to these temporal relations.

Consider first the case where *the stimulus-produced increment to drive does not appear until after a response has terminated*. Here it is obvious that the drive, though genuine, cannot act to energize or intensify the completed reaction. Indeed, if the interval between the response and the stimulus is relatively long, the increment in drive should have no effect upon either the vigor of the behavior or upon the probability of its recurrence. As is well known, long-delayed punishments are quite ineffective as deterrents to the performance of undesirable responses. Moreover, if the onset of the drive is delayed, its reduction must also be postponed, and the effectiveness of any reinforcing power possessed by that reduction will be reduced. It seems unlikely, therefore, that a response will either be learned or unlearned if it is followed after an appreciable time by both an increase and a decrease in drive. The possibility remains, of course, that learning may occur under the special conditions suggested by Miller and Dollard, where the decrease is substantially greater than the increase, and where the temporal interval between response and stimulation reduction is minimal. It is also possible that when the time interval is short, the drive increment may lead to the abandonment of the response. Such a "punishing" effect might be observed when the cues for the completed response are still present and when the drive-arousing stimulus evokes reactions which are incompatible with the completed one.

Finally, let us examine the situation in which *the stimulus-initiated increment to drive is produced while a response is in progress*. Since the response is actually occurring, the tendency to perform it must be greater than that for any other act. If drive is assumed to intensify whatever response is dominant at the moment, it follows that the ongoing reaction, as long as it is dominant, will be facilitated rather than disrupted.

Evidence to support this contention is found in studies by Muenzinger (17) and Drew (2), who report that electric shock administered after the choice point facilitates learning, and in Ullman's (27) experiment in which the rate of food consumption increased if rats were shocked while eating. Similarly, Miller (14) found that a momentary shock delivered to rats on the stem of an elevated T maze increased their speed of running and decreased their errors. His rats had been trained to run to food when hungry but were thoroughly satiated at the time shock was administered.

It thus appears that *a step-up in drive occurring during a response can exert a dynamogenic influence on that response, providing the eliciting cues are exercising a high degree of control over the behavior.* Moreover, this magnifying effect should appear on the very first occasion when the drive is heightened, i.e., before the drive-initiating stimulus has had any opportunity to accumulate either associative connections with the ongoing action or any secondary reinforcing attributes. Any stimulus capable of arousing a drive, however, may also be rich in its own potentialities for eliciting specific movements. If so, its presentation may upset the initial dominance relations among the reaction tendencies, and the response in progress may suffer interference. The anti-drive-reductionists seem to feel that a negative effect of this kind should be predicted by their opponents' theory on every occasion when increased drive occurs during a response.

A drive reductionist can admit, therefore, that genuine drive increments can occur after the initiation of a response and that the response may, nevertheless, persist or even become stronger. And he can admit this without having to abandon his special hypothesis as to the mechanism of reinforcement. He merely asserts that the ongoing behavior should be intensified except when it is overridden by competitive reactions. Though he assumes that a reduction in drive is reinforcing, he need not also assume that an increase in drive, in and of itself, produces an obverse negative effect.

In summary, we have been concerned in this discussion with some recent criticisms of the drive-reduction hypothesis of reinforcement. Though these criticisms differ in their manifest content, all are grounded in the observation that when a rise in stimulus intensity takes place during a response, or shortly thereafter, the probability that the response will recur does not always become less, and may even grow greater. If an increase in stimulation is alone sufficient to define heightened drive, then this observation constitutes an apparently crucial exception to the drive-reduction view, because the response is first followed by increased drive and only later by its decline. The fact that organisms often behave so as to obtain more stimulation rather than less, or to obtain so-called "pleasant" stimulation, constitutes, in essence, an identical apparent exception to the drive-reduction view.

In seeking to solve this problem, it has been denied that the drive state, if any, resulting from or induced by stimulation can be usefully defined solely in terms of the sheer physical intensity of the stimulus. The definition must be broadened to include such additional factors as amount and kind of previous experience with a stimulus, level of sensory adaptation, and degree and kind of simultaneous stimulation. When this is done, many of the supposed exceptions to the drive-reduction hypothesis are eliminated, since they can no longer be correctly described as involving increases in drive.

But even when all conditions of the modified definition have been met, some stimuli still qualify as drive arousers. The effects of these upon behavior can be coordinated with the drive-reduction view by contending that if they occur at appropriate times they will facilitate rather than disrupt behavior. Instances in which disruption does take place can be attributed to the elicitation of interfering responses by the drive-producing stimuli, and not to any negative motivational attributes inherent in them.

REFERENCES

1. Butler, R. A. Discrimination learning by rhesus monkeys to visual exploration motivation. *J. comp. physiol. Psychol.*, 1953, **46**, 95–98.
2. Drew, G. C. The function of punishment in learning. *J. genet. Psychol.*, 1938, **52**, 257–267.
3. Farber, I. E. Anxiety as a drive state. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln, Nebr.: Univer. of Nebraska Press, 1954. Pp. 1–46.
4. Gantt, W. H. Psychosexuality in animals. In P. H. Hoch & J. Zubin (Eds.), *Psychosexual development in health and disease.* New York: Grune & Stratton, 1948. Pp. 33–51.
5. Guthrie, E. R. *The psychology of learning.* New York: Harper, 1935.
6. Hardy, J. D., Wolff, H. G., & Goodell, Helen. Studies on pain. A new method for measuring pain threshold: observations on spatial summation of pain. *J. clin. Invest.*, 1940, **19**, 649–658.
7. Harlow, H. F. Motivation as a factor in the acquisition of new responses. In *Current theory and research in motivation.* Lincoln, Nebr.: Univer. of Nebraska Press, 1953. Pp. 24–49.
8. Hebb, D. O. *The organization of behavior.* New York: Wiley, 1949.
9. Hilgard, E. R. *Theories of learning.* New York: Appleton-Century-Crofts, 1948.
10. Keller, F. S., & Schoenfeld, W. N. *Principles of psychology.* New York: Appleton-Century-Crofts, 1950.
11. Liddell, H. S. Animal studies bearing on the problem of pain. *Psychosom. Med.*, 1944, **6**, 261–263.
12. Masserman, J. H. *Principles of dynamic psychiatry.* Philadelphia: Saunders, 1946.
13. McClelland, D. C. *Personality.* New York: William Sloane, 1951.
14. Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.
15. Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.
16. Mowrer, O. H. Two-factor learning theory: summary and comment. *Psychol. Rev.*, 1951, **58**, 350–354.
17. Muenzinger, K. F. Motivation in learning. I. Electric shock for correct response in the visual discrimination habit. *J. comp. Psychol.*, 1934, **17**, 267–277.
18. Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.
19. Parsons, C. M., & Goetzl, F. R. Effect of induced pain on pain threshold. *Proc. Soc. exp. Biol., N. Y.*, 1945, **60**, 327–329.
20. Pavlov, I. P. Conditioned reflexes. (Trans. by G. V. Anrep.) London: Oxford Univer. Press, 1927.
21. Peters, H. N. Experimental studies of the judgmental theory of feeling: I. Learning of positive and negative reactions as a determinant of affective judgments. *J. exp. Psychol.*, 1938, **23**, 1–25.
22. Postman, L. The history and present status of the law of effect. *Psychol. Bull.*, 1947, **44**, 489–563.
23. Skinner, B. F. *The behavior of organisms.* New York: D. Appleton-Century, 1938.
24. Slutskaya, M. M. [Converting defensive into food reflexes in oligophrenics and in normal children.] *Zh. Nevropatol.*, 1928, **21**, 195–210. (Reviewed in Razran, G. H. S. Conditioned reflexes in children. *Arch. Psychol.*, 1933, **23**, No. 148.)
25. Spragg, S. D. S. Morphine addiction in chimpanzees. *Comp. Psychol. Monogr.*, 1940, **15**, No. 7.
26. Stone, L. J., & Jenkins, W. L. Recent research in cutaneous sensitivity. I. Pain and temperature. *Psychol. Bull.*, 1940, **37**, 285–311.
27. Ullman, G. A. The experimental production and analysis of a "compulsive eating symptom" in rats. *J. comp. physiol. Psychol.*, 1951, **44**, 575–581.
28. Waters, R. H. The law of effect as a principle of learning. *Psychol. Bull.*, 1934, **31**, 408–425.
29. Wolff, H. G., & Goodell, Helen. The relation of attitude and suggestion to the perception of and reaction to pain. *Res. Publ. Ass. nerv. ment. Dis.*, 1943, **23**, 434–448.

(Received June 21, 1954)

Psychological Review
Vol. 62, No. 3, 1955

THE FUNCTIONS OF SCHEMATA IN PERCEIVING

M. D. VERNON
University of Reading, England

THE DEVELOPMENT OF SCHEMATA

There has been a striking tendency among psychologists in recent years to stress the importance of motivational factors in determining our perceptions of the outside world. Perception has even been regarded as "an approach to personality" (see Blake and Ramsey, **5**), as though a study of the individual's percepts would throw some light upon the nature of his individual personality characteristics. These psychologists appear to underestimate the importance of the cognitive aspects of perception; of the parts played by thought and reasoning in building up our understanding of the world; and of the functions of our background of knowledge[1] in determining our immediate percepts. At the same time, these psychologists seem to overestimate the influence of the factors of immediate need, and especially of the relatively transient and unimportant influences they exert upon most of the simpler and more straightforward of our perceptions. It is possible to hold the view that percepts are constructed from sensory data though not wholly a function of them; and at the same time to maintain that the construction is determined by cognitive activities as much as by individual need and interest. I shall proceed to discuss this view.

It is clear to all that at the basis of our perceptions of the external world there are a series of effects in the sense organs produced by external stimulation which vary in an orderly way, both temporally and spatially. And, as Gibson (12) has pointed out, our percepts vary also in gradients which correspond closely to the gradients of stimulation. Moreover, there is an orderly continuity of sensory impression between one moment and the next; and there is also a consistency and correspondence between the types and variations of effect in the different sense organs. From these patterns of sensory impressions are constructed our percepts of the external world. We assume that our percepts represent accurately what is "really there." We believe that certain visual impressions represent solid objects of particular shapes, which continue to exist indefinitely; and this probably because we can repeatedly check the fact of their solidity and identity by touching and manipulating them. For most practical purposes, tactile and kinesthetic data can be relied upon to be consistent.

It seems fairly clear from the evidence of Piaget and others (discussed in Vernon, **44**) that the infant is not at first aware either of the orderly repetition and continuity of sensory data, or of the association between such data from different sense organs. Neither does he allocate such impressions to sources in a more or less permanently existing external environment. He gradually acquires the processes of constructing his percepts of the external world through manipulation and experiment. It may well be that the tendency to do so is innate and develops through maturation; but the infant does require specific individual practice and experience before he arrives at the concept of the world around him as permanently existing outside himself and independent of himself. Indeed, the notion of the complete independence and objectivity of the external world seems not to be fully established until many years after birth.

---

[1] "Knowledge" is used throughout this paper in its widest sense, to include not only verbally formulated knowledge and ideas, but also our background of knowledge of how to behave appropriately in the widely varying range of situations we normally encounter in our everyday life.

Now is it not obvious that this view of the external world must be constructed by the individual for himself, since it cannot be directly "given" by the sensory impressions coming from the sense organs? And since, sooner or later, our perceptions and ideas about the world do become in general coherent and consistent, is it not probable that they are coordinated in accordance with some system of classification of percepts and ideas about them? If the infant is to be impressed by the repetition, coherence, and consistency of certain combinations of sensory data, he must adopt some organized procedure of retaining, coordinating, and classifying these data in an orderly manner. The earliest classifications of data may be formed in relation to the satisfaction of the needs common to all children; they center upon the mother, who is the supplier of food, warmth, and comfort. But, also, the infant and the young child are intensely interested in the appearance of the phenomena they experience, and make every effort to study and understand them. Thus data are classified in terms of the appearance and behavior of external objects and situations. These classifications supplement and are substituted for the classifications in terms of need satisfaction. As language develops, it forms a tool of increasing value in assisting the former type of classification.

I suggest that to this type of classification we may usefully apply Bartlett's concept of the "schema" (3). Bartlett defines the schema [2] as follows:

> "Schema" refers to an active organization of past reactions or of past experiences, which must always be supposed to be operating in any well-adapted organic response. That is, whenever there is any order or regularity of behavior, a particular response is possible only because it is related to other similar responses which have been serially organized, yet which operate, not simply as individual members coming one after another, but as a unitary mass (3, p. 201).

Thus we may regard schemata as persistent, deep-rooted, and well-organized classifications of ways of perceiving, thinking, and behaving. They are based upon the individual's knowledge of "what sort of thing to do in, and about, certain kinds of situation." Thus they involve the classification of (*a*) situations of related kinds; and (*b*) kinds of behavior appropriate in these situations. But the behavior and its results react back upon the classification of situations. If the results prove unsatisfactory, the classifications of both situations and behavior may be modified until they appear to be more realistic and more appropriate. Thus schemata are not in themselves stereotyped habit systems, though in certain circumstances it may be that they will degenerate into such habit systems.[3] There is no doubt, however, that the individual's motivational tendencies will operate in such a way as to build up and strengthen schemata which lead to behavior that satisfies need and to modify or suppress those which do not.

The essential point to note is that the use of a concept such as that of the schema reflects the fact that our percepts, thoughts, and behavior are on the whole consistent and orderly—reasonably appropriate though not necessarily logical—and that they are consistently related to previous acts of perceiving, thinking, and behaving. Thus, in investigating the operation of schemata, we must endeavor to find if the observer shows some consistent method of classifying perceptual situations in a realistic and meaningful manner. Furthermore, we must note if there is evidence that this type of classification has been built up through individual personal experience or interest, in which case there may be considerable individual differences in perceiving the same situation (such as the responses I have quoted elsewhere [44, pp. 262–263] to a reproduction of Sickert's picture, "Brighton Pierrots"). Or if the systems of classification are shared by members of a group having, for instance, a common culture or type of education, we may expect that the percepts of these people will be alike but will differ from those of other groups. Lastly, there are many schemata related to the spatial and temporal relationships of objects in the external environment which are probably acquired by almost everyone in the course of his ordinary everyday life. It is not always easy to detect the manner in which perception is determined schematically, since often we are unaware of the background of knowledge and experience to which our present perceptual behavior is related. But it is, I think, quite essential to the understanding of our immediate perceptual reactions to take into account these organized classifications of knowledge about the physical and social environment which guide and inform us as to "what is the sort of thing to do about the sort of thing we now see."[4]

---

[2] For a fuller discussion of this definition, and of its relation to other uses of the term "schema," see Oldfield and Zangwill (26).

[3] Recently H. Kay has, in an unpublished thesis, given evidence that old people's behavior shows signs of such deterioration.

OPERATION OF SCHEMATA IN EXPERIMENTS ON PERCEPTUAL AND FORM QUALITIES

Although the sensory data are accurately recorded at some level in the central nervous system, nevertheless, it does not seem that we are ever directly or consciously aware of anything corresponding to the sensations of Titchenerian theory. Instead, we perceive stable solid objects, enduringly located in the world around us. This "visual world" is perceived, and we react to it, in accordance with the schemata we have built up for ourselves from infancy upwards. But we do also perceive that certain aspects of the external world do at times vary independently of one another, and that they cannot, and must not, always be determined by our ideas of the general nature of the external world and the objects which it contains. Perhaps only the sophisticated Westerner has gone to much trouble to distinguish these particular aspects and their inherent differences and variations. For it is only by a considerable effort of differentiation and abstraction from his total perceptual experience that he can perceive them at all. Moreover, this abstraction is never complete. What is perceived is always affected to some extent by the global schematized perception of the total field; and other aspects of the field are always liable to slip in and distract the observer.

Now it seems that most of the *experimental* work which has been carried out on perception depends on the ability of the observer to abstract more or less completely certain aspects of the perceived field from the total global impression. In particular, all so-called "sensory" experiments are of this nature. They indicate that the trained and sophisticated observer can analyze his perceptual field, and differentiate certain parts of it from the remainder, with considerable finesse. Furthermore, it seems to be true that the degree of possible differentiation is fairly consistent, both in the same observer from time to time, and also between different observers with the same degree of training.

---

[4] This appears to correspond to Postman's views (31) on the categorization of "perceptual response dispositions."

It has been assumed that in such experiments, the abilities of the sense organs to respond to and discriminate stimuli are being measured. It would be more correct to say that when the situation is arranged in such a way that the physical stimulus is varying in one way only, and all other stimulus qualities are maintained constant, then the variations in what is perceived are largely a function of that stimulus variable. But appropriate schemata are nevertheless operating in the construction of the percepts. For they will vary greatly between different observers with different degrees of sophistication and training, and those who have built up appropriate schemata of observation and response will react differently from those who have not. Furthermore, in situations which are not limited in this way, other schemata will operate, and the percepts will be differently constructed from the sensory impressions.

Again, in many experiments on perception, as Gibson (13) has stressed, the observer is required to abstract from his total experience small differences in the outlines of two-dimensionally projected shapes, shown to him momentarily. He learns by experience to construct some percepts of this type with considerable facility. Whether all children learn to recognize representations of objects from their outlines is impossible to say. Certainly children brought up in Western civilizations acquire this facility at an early age, presumably by coordinating together visual impressions of brightness gradients and tactually perceived "edges." But when it is a matter of perceiving nonrepresentational outline drawings, of the kind so frequently used by the gestalt psychologists, the individual has little previously acquired experience to assist him. Then his percepts may be characterized by the failure to differentiate the parts from the whole, as is shown in many of the gestalt figures and in the visual illusions.

OPERATION OF SCHEMATA IN "CONSTANCY" EXPERIMENTS

In experiments on the "constancies," the observer is required to differentiate and abstract aspects of his total experience in a way that is quite unfamiliar to him. The child has learned in the course of his development that objects may retain their identity, and hence their "real" shape and size, although the visual aspects of their shape and size vary as their spatial position varies. This he learns at a comparatively early age with objects which he can reach and handle, but at a much later age, as Piaget (27, 28) has shown, with distant objects which he cannot manipulate. In the end, however, he can make this act of identification even when, for instance, distant objects appear actually as little more than dots on the landscape. Furthermore, on the basis of impressions of projected size and shape, he can and does make fairly accurate estimates of the distance and spatial orientation of the objects. What he is never required to do, and cannot do with any accuracy, under normal conditions, is to estimate their projected size and shape. (Piaget and Lambercier [30] have shown that this is done more accurately by younger children, aged 7–8, whose percepts are less closely linked to the spatial framework, than by older children, aged 9–10; after that, the ability improves somewhat as they become more capable of making abstract judgments.) Yet this is what the observer is required to do in experiments on size and shape constancy. The unsophisticated observer naturally finds this task difficult. In estimating or matching the size of distant objects, he is apt to choose the size which the object would appear to have if it were close to him. He uses both his knowledge of the size of the object as such, and also schematized knowledge relating this size to the retinal projections of objects at distances which can be judged from other criteria. But suppose that, instead of being required to estimate the perceived size of the object, he is asked to estimate what fraction or segment of the total field of view it covers. Then, as Joynson (17) has shown, the estimated size approaches that of the retinal projection of the object. Much the same procedure must be adopted by anyone who attempts to make a landscape drawing. He must try to estimate the size of the spaces between objects, as well as of the objects themselves; otherwise the latter will be drawn too big to fit into his drawing. Thus the object is judged to be of one size if it is regarded "figurally," as an object in itself, and of another size if it is perceived as part of the "ground."

A somewhat different situation seems to arise in experiments on shape constancy. Here the observer rarely judges the shape of the inclined object to be exactly the same as that of the object in the frontal parallel position, unless the angle of inclination is very small. Thus if the object is a circle, his judgments become more and more elliptical as the angle of inclination is increased. It is possible that in the course of our upbringing we learn to perceive two-dimensional shape as such, and apart from the object concerned, with a fair amount of accuracy. The observer attempts, in the experiment, to apply this schematic perceptual category to the estimation of projected three-dimensional shapes. In effect, he says to himself, "This tilted circle looks elliptical." But he cannot judge very accurately how elliptical.

It is not justifiable, I consider, to compare the processes of estimation which occur under normal conditions with those which operate when the object is exposed in such a way that only its projected shape is visible. Such estimates are made when the object is viewed monocularly, with no variations of lighting or surface texture, and with all surroundings concealed. The normal schemata are not brought into action, and the variations of the percept are determined by the only variation perceptible—that of projected shape. In the experiments of Ittelson (15), however, the observer perceived a single shape continually varying in size. His schematized ideas as to the persistent nature of objects, which frequently change their distance but not their size, suggested to him that, when the area of retinal stimulation varied, the object was in fact moving rather than expanding or contracting; and that was what he perceived. Again, in the experiments of Langdon (20), a circular outline shape made of wire was rotated about a vertical axis, in such a way that its projected shape varied continuously. The shape of the object was then estimated as somewhere between that of the circle and the projected shape, thus approaching the estimate that would have been made under normal conditions of perception. Possibly, once more, the continuity of experience was the important factor; and the rotating outline was perceived in somewhat the same way as a stationary solid object whose spatial orientation was clearly visible.[5]

A rather different schematic determination of perception again appears to operate in brightness—or "whiteness"[6]—and in color constancy. Here also it is true that what we are concerned with in ordinary everyday life is the intrinsic or normal whitenesses and colors of objects—those which appear in ordinary daylight. Over a wide range of brightness and hue in general illumination, we do ignore changes in perceived whiteness and color of objects. At the same time, if an observer is asked to match the whiteness or color of white objects exposed in dim or colored light, against gray or colored objects exposed in bright white light, it will appear that his perceptions of the former have become appreciably modified—that is to say, they are less white or are tinted with color. Now it may be that although we can accurately differentiate between two whitenesses or two colors placed side by side, we have no schematic categories appropriate to the remembering of absolute degrees of whiteness or color. Anyone who has tried to remember the exact hue of a fabric in order to match it against another, even in white light, will confirm the difficulty of this task. We have, it is true, developed some schematic knowledge of the changes in whiteness and color of objects with changes in brightness and hue of the light falling upon them. It may be, as Wallach argues (47), that we have learned to assess the proportion of incident light which is reflected to us, rather than the absolute amount. Thus whiteness constancy is perceived as long as the proportion or ratio remains approximately constant. But this judgment is not completely accurate, and constancy of whiteness and color is in general less than constancy of shape and size.

---

[5] Langdon states that it often appeared to the observers that the wire shape was being squeezed in and out, rather than rotated, but that this made no difference to their estimates. It is possible that they developed a procedural "set" in matching which was too well-established to be disrupted.

[6] In accordance with the usage of Thouless (42) and of Koffka (18).

I do not of course claim that this discussion of the "constancies" has covered exhaustively all the phenomena they exhibit. This argument merely advances a new approach to the subject, which might suggest a somewhat different type of experimental investigation, for instance, studies of the relation of constancy judgments made in different types of naturalistic surroundings, and with different types of experimental instructions. Those of Gibson (12) constitute a valuable beginning which needs to be followed up. Above all, it is important that the observers should attempt to report in some detail the nature of the phenomena as they actually perceive them, and thus demonstrate the individual differences in their perceptions.

RELATION OF SCHEMATA TO MOTIVATIONAL FACTORS

I have argued that we perceive certain aspects of the perceptual field in accordance with the schematic category of events to which at the moment they seem to appertain. Since these schemata develop differently according to different individual experience, it cannot be assumed that percepts which may be claimed to be, logically and objectively, of the same kind—for instance, the "constancies"—will in fact follow the same logical principles or "laws." Furthermore, the operative schemata can up to a point be varied by the intentions of the observer—by his concern with certain aspects of the field, and his lack of interest in others. The individual can choose to some extent what he will perceive. But it has generally been found that in carrying out skilled perceptual tasks, for instance, the observer cannot perform satisfactorily, however strong his motivation, merely by intending from moment to moment to concentrate on particular aspects of the field. He must be instructed beforehand where and how to direct his awareness, and he must acquire facility in this by repeated practice. In other words, he must build up the appropriate schemata for the selection, assimilation, and interpretation of the appropriate sensory data. The formation of the schemata is strongly motivated and stimulated by interest, but the actual percepts are not directly determined by the motivation. The construction of such schemata is a necessary part of the training of aircraft spotters, radar and Asdic operators; and in a variety of industrial tasks of inspection and of grading materials.

The schemata operating in perception perform two functions: (*a*) They produce a condition of expectation in which the observer is not merely on the *qui vive,* but also knows what to look for—what particular sensory data to select from the incoming flood. (*b*) He then knows how to deal with these data—how to classify, understand, and name them, and draw from them the inferences that give the meaning to the percepts. We are sufficiently familiar with the case of the unsophisticated student who does not even see the minute details of, for instance, biological specimens, from which the expert can derive the most important information. An interesting study of such cases by Johnson (16) demonstrated the manner in which students could be directed to observe the essential points of specimens, etc. seen under the microscope, which they had hitherto overlooked, and not to make incorrect inferences about them. Another instructive example of this phenomenon occurred in an experiment by Fox (11), who showed pictures of armor to students, and required them afterwards to report what they remembered of the pictures. He found that if he gave the students some preliminary instruction on technical details, and if they thoroughly understood and assimilated this information, then subsequent performance was superior to that of the uninstructed. But if, on the other hand, they did not fully assimilate the instruction—that is to say, it had not been fully absorbed and schematized—then it did not assist them at all, but merely confused them.

Now it might be argued that such highly skilled perceptual tasks were characteristic only of the expert, and had little relationship to the perceptions of ordinary everyday life. The distinction seems to be one of degree rather than of kind. It was pointed out above that young children were, in general, intensely interested in and curious about the external world and the objects and events within it. Is it not probable that for many of them the same features of the environment are important and interesting, and that therefore they build up similar schemata on which to base their subsequent perceptions? Thus Piaget has described the development of schematized ideas and corresponding perceptions of movement (29), causal phenomena (27), distance and space (28), and many others. Though his observations covered only small numbers of children in one particular environment, it seems probable that all children must develop such schemata sooner or later in order to adjust themselves to the surrounding world. But insofar as their environments differ, for instance, as between town and country dwellers, then different schemata are also developed.

Such differences in individual experience also seem to operate in what is called "social perception." It is a subject too wide to discuss here in detail. But I would include in it both the ability to perceive and act upon fine shades of behavior and emotional expression in others; and also differences in the construction of percepts from given sensory characteristics. There is little doubt that particular interests and sentiments affect the ability to develop schemata for classifying and responding appropriately to what we perceive in the behavior of others. As regards the selection of particular sensory data, it appeared in the experiments of Marks (24) that individuals with pro-Negro sentiments had developed schemata which enabled them to perceive and remember in greater detail than those without those sentiments the characteristic facial differences between different Negroes. In the experiment of Seeleman (41), Negroes themselves seemed to have established schematized scales of facial color, which in part determined the estimates of color which they made in the experiment.

It is possible that some long-term schema based upon the interests of a particular social class was operating to cause the overestimation of size of the more valuable coins in the famous experiment of Bruner and Goodman (6). However, the dust of controversy has not settled upon the methods and results of this experiment. It seems fairly certain that, as in the experiments described by Ansbacher (1) and Postman and Page (37), there was an inability to isolate a single aspect from a global percept. But one would need to know more about the poor children to decide why their overestimation was greater than that of the well-to-do.

However, in these and other cases the schemata were very complex, and the factors which brought different schemata into operation from moment to moment are obscure. Although it seems probable that in prolonged and continuous perceptual activity, highly developed and motivated schemata will tend to operate, it can by no means be assumed that any particular schema will function at any particular moment, particularly in somewhat ambiguous and indeterminate situations. In some experiments, such as those on tachistoscopic perception, there may be nothing to inform the observer as to which, if any, of his schemata are appropriate to the particular material presented to him. In such circumstances, schematic determination of the perceptual response may be weak and unclear; no one schema can function very effectively, and the observer may be quite uncertain what to look for and to select from the visual field. He is then prone to develop, as a reaction to the experimental situation, a cognitive "set" which causes him to expect and hence to perceive the type of material indicated by the instructions or the conditions of the experiment. Thus in one experiment (43), I presented tachistoscopically to a number of airmen a series of rather bizarre pictures, terminating in a simple picture of an airman in a cockpit. Several of them failed to recognize the picture, and it seemed that this failure was due to a set established by the preceding pictures to perceive "something queer." In this situation they had no reason to suppose that any particular schema, for instance one related to their professional interests, would be appropriate; thus their perceptions were determined by the short-term cognitive set. Of course, had this type of situation been indefinitely repeated, and had they been satisfied that their responses were appropriate to the situation, the set might in time have developed into a schema related to this situation. Such schemata, as was noted above, may be built up by observers who have had prolonged training and experience in experiments on sensory qualities.

We may now proceed to examine some of the experiments which have purported to demonstrate the direct effects of short-term need states upon perception. It must be noted that in many of these experiments, also, there may have been no schema available to direct, in a clear and unambiguous manner, the observer's expectations of what would appear, or his selection from the sensory data presented. But, indeed, in some of these experiments it hardly seems that we are dealing with perception based upon external stimulation; but rather with the projection of internal imagery upon the external situation. The latter process appears to have occurred in the experiments of Levine, Chein, and Murphy (21) and of McClelland and Atkinson (22) on the reporting of "food objects" by hungry observers in response to subliminal stimulation, or to no stimulation at all. Yet Postman and Crutchfield (36) have since shown that even a temporary cognitive set established by instructions to expect to see words connected with food was more effective in stimulating the perception of such words than was actual hunger. After all, why should a hungry individual expect to see words connected with food, or even pictures of food, in a tachistoscope? Why should he wish to do so? But there is nothing to be gained by confusing processes which are initiated by sensory stimulation with those for which there is no known sensory basis. This applies also to the Rorschach responses, for instance, which consist of imagery that the observer himself can differentiate from the immediate percepts (cf. data reported in the chapter by Dennis in Blake and Ramsey, 5). Indeed, it is probably true of all situations in which the sensory data are so scanty that the individual is forced to construct something imaginatively. Even if it is impossible to draw a hard and fast line between these situations and those of normal everyday perception, it is still undesirable to extend conclusions based upon the former to cover the latter also.

In many of the genuine perceptual experiments employing the perception of words, it appears that perception was determined by the long-term schemata which direct this extremely subtle and highly skilled activity. In particular, the effect of familiarity, a function of these schemata, is always great. No accurate compensation for individual differences in familiarity can be provided by the use of the Thorndike-Lorge word count, since highly educated individuals vary greatly in their familiarity with the more uncommon words, and it is impossible to measure this variation. Familiarity, however, is probably at least in part a function of schemata built up in accordance with individual interests. The individual develops particular sets of ideas, and words to express them, which are related to his interests. This indeed appeared in the experiments of Postman, Bruner, and McGinnies (35) and of Postman and Schneider (38). Here the tendency was for observers to perceive most readily words, and especially unusual words, which were related to their prevailing interests.

In another type of experiment, for instance, those of Verville (45) and of Verville and Cameron (46), the observer's ability to perceive incomplete pictures appeared to be inhibited as the result of certain frustrating experiences which they had received not long before. The same effect upon the reading of words was demonstrated by Postman and Bruner (34). It seems probable that in these experiments the observer became accustomed to the occurrence of a task which was too difficult for him to accomplish, and in most cases resigned himself to failure. However, Eriksen (10) reported that observers who forgot their previous failures were slower than those who did not, to perceive words which were likely to affect them emotionally—those to which they gave a long reaction time in a free association test. Thus it may be that in some cases an experience of failure may stimulate the observer and set up an increased concentration of attention.

But most of the experiments on the so-called "perceptual defense" have probably been affected by differences in familiarity with the "taboo" words presented. Again the Thorndike-Lorge word count seems singularly inadequate for equating the familiarity of these words with the familiarity of the "neutral" control words. Familiar acquaintance with "rude" words, the readiness to report them without hesitation, and the whole complex of emotions tied up with them, are likely to vary enormously among different individuals. Indeed, considerable variability in these respects was indicated in the experiments of Bruner and Postman (7) and of Cowen and Beier (8); though Bruner and Postman provided a control on individual differences by measuring the reaction times in free association. That many of the observers did dislike saying the taboo words out loud was clearly shown in the comments quoted by Bitterman and Kniffin (4) of observers whose susceptibilities had obviously been outraged! However, it seems that this effect wore off as the observers became accustomed to the situation, according to the results of Lacy, Lewinger, and Adamson (19). That it could be removed by allowing the observers to write the words, and by giving them special instructions to do so, appeared from the experiment of Postman and Gropper (32). Once these difficulties had been overcome it seemed that, in general, the taboo words were perceived *more* easily than the neutral words, rather than more slowly, as McGinnies (23) originally suggested. Thus it may be that in certain circumstances the ease of perception may be increased by emotional excitement. However, the conclusions drawn from these experiments are of doubtful import, because of the effects of differences in familiarity; and the same may also be true of experiments carried out with pictures, such as those of Eriksen (9). It would be difficult to conclude that these emotional factors determined the actual nature of what was perceived.

We come now to experiments in which it did appear that what the observers perceived was in part a function of motivation set up during the experiment. In the experiments of Proshansky and Murphy (39) and of Schafer and Murphy (40) the observers were found to accentuate certain aspects of what they perceived and to reject others, insofar as the former were presented in conjunction with a "reward," and the latter in conjunction with a "punishment." For the reward money was given them; and for the punishment the money was taken away from them. The experimenters satisfied themselves that the observers did not become consciously aware of the connection between the particular aspects of the field, and the reward or punishment, so that they did not deliberately choose the former and reject the latter. But it is important to note that in the experiment of Schafer and Murphy, the rewarded aspect was favored by contrast with the punished only over a limited period. After about sixteen responses, the observers developed cognitive sets based upon the nature of the experimental material, which thereafter completely determined what they perceived and eliminated the effects of the reward and punishment. Thus, in these experiments the reward and punishment situation did for a time affect the observers' selection of certain aspects of the field. But the effects were apparently unstable; and there was a tendency for the observer to develop a cognitive set which was more stable and effective. This result parallels that of Postman and Crutchfield (36) with the food words. Haggard and Rose (14) tried to show that the amount of additional activity carried out by the observer in connection with his perceptual activity was also a factor of importance; and it may be that such activity would be related to the establishment of sets. But the evidence was obtained from experiments on the "autokinetic effect," a phenomenon which cannot be termed wholly perceptual since it is at least partly illusory.

Stronger evidence of the direct effect of motivation upon perception is given by the experiment of Postman and Brown (33). They found that observers who had been told that they had "succeeded" (surpassed their level of aspiration) in certain tasks were relatively quicker than those told they had "failed" (fallen short of their level of aspiration) to perceive words related to success (excellent, succeed, perfection, winner); whereas those told they had failed were quicker than those told they had succeeded to perceive words related to failure (failure, unable, defeat, obstacle). Here it seems unlikely that the results could be attributed to differences in familiarity with the different words. Moreover, the experiences of success and failure were set up in the experimental situation and did not reflect any inherent individual tendencies to optimism or defeatism. There is at present no evidence as to how this effect upon ease of reading these particular words would compare in extent or permanence with that of a cognitive set or established schema. But the motivation had at least a temporary effect upon the observer's selection from the sensory data supplied to him.

The fundamental criticism of the experiments designed to show the direct effects of motivation upon perception is that they do not appear to take into account what this motivation might make the observer expect to see. As we have noted, in situations closely linked to an established interest, and those in which the observer received some specific training, he is on the lookout for certain aspects of the field of view and thus perceives them readily. But in the experimental situations, he may not know what to expect; and he has no reason, in most cases, to expect that anything related to the satisfaction of his needs will appear. Thus sets are readily established by the experimental conditions to perceive certain phenomena which are stressed by these conditions. To determine more clearly the effect of motivation, further thought and experiment are required to work out what are the aspects of a complex perceptual field which an observer in a state of unsatisfied need might look for and select. It may then appear that those features are the ones which have become schematically related to the need state and to the behavioral responses to the field which tend to satisfy the need.

CONCLUSIONS

It is true that we cannot at the moment predict the exact mode of operation of the schemata of any given individual in any particular set of circumstances. Neither can we hypothesize as to their neurological basis; although Oldfield (25) has recently advanced a suggestion as to how they might be explained in terms of the circuital storage devices of modern computing machines. But even in default of such an explanation, I would argue that the concept of schemata and their functions is of value in systematizing the following observations:

1. Percepts of the surrounding world, the objects it contains and their interrelationships, are not infinitely variable, sporadic, and inconsistent. Neither do the percepts vary in a one-to-one relationship with the sensory impression produced by external stimulation, although there is obviously a close correspondence.

2. Percepts are consistent; they have a continuity of meaning and make sense. The meaning and the sense are established by the coordinated knowledge of the observer about the external world and its happenings, and about the types

of behavior appropriate and satisfactory in different perceived situations.

3. The percepts of different observers of the same situation may in certain circumstances differ greatly from one another. But these variations do not appear to be purely random ones; there is some correspondence between the modes of perceiving of any given observer. Different observers have built up different schemata in accordance with their differing individual experience and the interests which lead them to seek this experience. Their expectations and their selection of what they perceive are in part determined schematically. But far more experimental evidence is required to work out these relationships in detail; evidence which must be obtained by a careful study of individual cases, their particular knowledge and interests.

4. In many of the experiments on perception, the observer is presented with a situation in which his existent schemata cannot operate freely, nor tell him what to look for. The sensory data are transient or ambiguous, and cannot be readily fitted into existing categories of experience. In these circumstances, the observer is rather readily influenced by short-term cognitive sets based upon the actual conditions of the experiment. In suitable conditions, such sets may ultimately develop into long-term schemata. His perceptions may also be affected by temporary need states, though I have argued that these effects have not been as clearly established as some experimenters have maintained. Particularly, we have little evidence as to what an observer in such states of need might look for or wish to perceive in any particular experimental situation. It is therefore difficult to draw any conclusions from these experiments which throw light on the fundamental relations between perception and motivation. In contrast, in the perceptions of everyday life, schemata established by previous experiences are available to direct us in perceiving certain aspects of the environment and reacting to them appropriately. It would, therefore, be most valuable to study the consistent modes of perceptual behavior which are shown by different individuals in situations more like those of everyday life than the situations encountered in so many perceptual experiments.

REFERENCES

1. ANSBACHER, H. Perception of number as affected by the monetary value of objects. *Arch. Psychol., N. Y.*, 1937, No. 215.
2. ASCH, S. E. *Social psychology.* New York: Prentice-Hall, 1952.
3. BARTLETT, F. C. *Remembering.* Cambridge, Eng.: Cambridge Univer. Press, 1932.
4. BITTERMAN, M. E., & KNIFFIN, C. W. Manifest anxiety and perceptual defense. *J. abnorm. soc. Psychol.*, 1953, **48**, 248–252.
5. BLAKE, R. R., & RAMSEY, G. V. *Perception: an approach to personality.* New York: Ronald, 1951.
6. BRUNER, J. S., & GOODMAN, C. C. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33–44.
7. BRUNER, J. S., & POSTMAN, L. Emotional selectivity in perception and reaction. *J. Pers.*, 1947, **16**, 69–77.
8. COWEN, E. L., & BEIER, E. G. The influence of threat expectancy on perception. *J. Pers.*, 1950, **19**, 85–94.
9. ERIKSEN, C. W. Perceptual defense as a function of unacceptable needs. *J. abnorm. soc. Psychol.*, 1951, **46**, 557–564.
10. ERIKSEN, C. W. Defense against ego-threat in memory and perception. *J. abnorm. soc. Psychol.*, 1952, **47**, 230–235.
11. FOX, C. A study in pre-perception. *Brit. J. Psychol.*, 1924, **15**, 1–16.
12. GIBSON, J. J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950.
13. GIBSON, J. J. What is form? *Psychol. Rev.*, 1951, **58**, 403–412.
14. HAGGARD, E. A., & ROSE, G. J. Some effects of mental set and active participation in the conditioning of the autokinetic phenomenon. *J. exp. Psychol.*, 1944, **34**, 45–59.

15. ITTELSON, W. H. Size as a cue to distance: radial motion. *Amer. J. Psychol.*, 1951, **64**, 188–202.

16. JOHNSON, M. L. Discussion methods in pre-clinical teaching. *Lancet*, 1950, **259**, 313–317.

17. JOYNSON, R. B. The problem of size and distance. *Quart. J. exp. Psychol.*, 1949, **1**, 119–135.

18. KOFFKA, K. *Principles of Gestalt psychology*. London: Kegan Paul, 1935.

19. LACY, O. W., LEWINGER, N., & ADAMSON, J. F. Foreknowledge as a factor affecting perceptual defense and alertness. *J. exp. Psychol.*, 1953, **45**, 169–174.

20. LANGDON, J. Further studies in the perception of a changing shape. *Quart. J. exp. Psychol.*, 1953, **5**, 89–107.

21. LEVINE, R., CHEIN, I., & MURPHY, G. The relation of intensity of a need to the amount of perceptual distortion. *J. Psychol.*, 1942, **13**, 283–293.

22. MCCLELLAND, D. C., & ATKINSON, J. W. The projective expression of needs. I. The effect of different intensities of the hunger drive on perception. *J. Psychol.*, 1948, **25**, 205–222.

23. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.

24. MARKS, E. S. Skin color judgments of Negro college students. *J. abnorm. soc. Psychol.*, 1943, **38**, 370–376.

25. OLDFIELD, R. C. Memory mechanisms and the theory of schemata. *Brit. J. Psychol.*, 1954, **45**, 14–23.

26. OLDFIELD, R. C., & ZANGWILL, O. L. Head's concept of the schema and its application in contemporary British psychology. Part I. Head's concept of the schema. Part II. Critical analysis of Head's theory. Part III. Bartlett's theory of memory. Part IV. Wolters' theory of thinking. *Brit. J. Psychol.*, 1942, **32**, 267–286; **33**, 58–64; 113–129; 1943, **33**, 143–149.

27. PIAGET, J. *La construction du réel chez l'enfant*. Neuchâtel: Delachaux et Niestlé, 1937.

28. PIAGET, J. *La formation du symbole chez l'enfant*. Neuchâtel: Delachaux et Niestlé, 1945.

29. PIAGET, J. *Les notions de mouvement et de la vitesse chez l'enfant*. Paris: Presses Universitaires de France, 1946.

30. PIAGET, J., & LAMBERCIER, M. La comparaison des grandeurs projectives chez l'enfant et chez l'adulte. *Arch. de Psychol.*, 1951, **33**, 81–130.

31. POSTMAN, L. Perception, motivation and behavior. *J. Pers.*, 1953, **22**, 17–31.

32. POSTMAN, L., BRONSON, W. C., & GROPPER, G. L. Is there a mechanism of perceptual defense? *J. abnorm. soc. Psychol.*, 1953, **48**, 215–224.

33. POSTMAN, L., & BROWN, D. R. The perceptual consequences of success and failure. *J. abnorm soc. Psychol.*, 1952, **47**, 213–221.

34. POSTMAN, L., & BRUNER, J. S. Perception under stress. *Psychol. Rev.*, 1948, **55**, 314–323.

35. POSTMAN, L., BRUNER, J. S., & MCGINNIES, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, **43**, 142–154.

36. POSTMAN, L., & CRUTCHFIELD, R. S. The interaction of need set and stimulus structure in a cognitive task. *Amer. J. Psychol.*, 1952, **65**, 196–217.

37. POSTMAN, L., & PAGE, R. Retroactive inhibition and psychophysical judgment. *Amer. J. Psychol.*, 1947, **60**, 367–377.

38. POSTMAN, L., & SCHNEIDER, B. H. Personal values, visual recognition and recall. *Psychol. Rev.*, 1951, **58**, 271–284.

39. PROSHANSKY, H., & MURPHY, G. The effects of reward and punishment on perception. *J. Psychol.*, 1942, **13**, 295–305.

40. SCHAFER, R., & MURPHY, G. The role of autism in a visual figure-ground relationship. *J. exp.. Psychol.*, 1943, **32**, 335–343.

41. SEELEMAN, V. The influence of attitude upon the remembering of pictorial material. *Arch. Psychol., N. Y.*, 1940, No. 258.

42. THOULESS, R. H. Individual differences in phenomenal regression. *Brit. J. Psychol.*, 1932, **22**, 216–241.

43. VERNON, M. D. Different types of perceptual ability. *Brit. J. Psychol.*, 1947, **38**, 79–89.

44. VERNON, M. D. *A further study of visual perception*. New York: Cambridge Univer. Press, 1952.

45. VERVILLE, E. The effect of emotional and motivational sets on the perception of incomplete pictures. *J. genet. Psychol.*, 1946, **69**, 133–145.

46. VERVILLE, E., & CAMERON, N. Age and sex differences in the perception of incomplete pictures by adults. *J. genet. Psychol.*, 1946, **68**, 149–157.

47. WALLACH, H. Brightness constancy and the nature of achromatic colors. *J. exp. Psychol.*, 1948, **38**, 310–324.

(Received May 20, 1954)

Psychological Review
Vol. 62, No. 3, 1955

SYMPOSIUM ON THE PROBABILITY APPROACH IN PSYCHOLOGY *

REPRESENTATIVE DESIGN AND PROBABILISTIC THEORY IN A FUNCTIONAL PSYCHOLOGY [1]

EGON BRUNSWIK

University of California

The movement in psychology most directly concerned with the challenge of unification with the exact sciences is behaviorism. At the present time, two major factions are discerned within behaviorism. One, represented chiefly by Hull (26), tries to elaborate classical behaviorism into a tightly woven, formalized system of strict laws about intraorganismic processes in the nomothetic tradition of physics. The other, inaugurated chiefly by Tolman and his program of molar or "purposive" behaviorism (47), attempts to deal with behavior "relative to some end" and thus to restructure behaviorism in the tradition of the utilitarianism of Darwin and of early American functionalism.

It may be argued that nomothetic behaviorism overexpands physicalism beyond the necessary observational and procedural core and includes unessential borrowings from the specific thema of physics. A functionally oriented objective psychology, on the other hand, dealing as it does with organism-environment relationships at the more complex level of adjustment, may be seen as falling in line with a more searching interpretation of the historical mission of psychology. The present paper concentrates on summarizing and expanding earlier contentions (2, 7, 8, 48) to the effect that the environment to which the organism must adjust presents itself as semierratic and that therefore all functional psychology is inherently probabilistic, demands a "representative" research design of its own, and leads to a special type of high-complexity, descriptive theory.[2] This program provides not only the necessary thematic diversification from the classical

---

* The series of papers by Egon Brunswik and by Leo Postman, and of the following discussion papers by Ernest R. Hilgard, David Krech, Herbert Feigl, and Egon Brunswik, are adapted from the first part of a symposium held, under the same title and under the chairmanship of Edward C. Tolman, at the *Berkeley Conference for the Unity of Science*, University of California, July 1953.

A contribution by Kenneth R. Hammond to a second, more practically oriented part of the original symposium will appear in the next issue of this journal (20).

The Institute for the Unity of Science in Boston, which sponsored the conference, has contributed approximately one-half the publication cost of these papers.

[1] The present paper has been considerably expanded beyond the original exposition read at the symposium of which it was part. However, care was taken not to alter the substance of the argument on which the subsequent paper by Postman and the ensuing discussion are based.

[2] The expansions beyond the earlier publications listed concern mainly the use of a behavioral example at the beginning of the paper; the brief consideration of such semirepresentative policies as "canvassing"; certain comparisons with factorial design and the analysis of variance, as well as with nonfunctionalistic uses of probability in psychology; and a discussion of actual and potential applications to the clinical-social area and to related domains.

natural sciences but also leads to the long overdue internal unification of psychology.

We will develop our arguments first with the use of an example involving behavior as a "constant function" of a characteristic end state, and then in reference to the functionalism of the perceptual constancies where progress along methodological lines is somewhat further advanced.

SYSTEMATIC DESIGNS AND THE STUDY OF DISTAL ACHIEVEMENT

Behavior as constant function. One of the earlier functional behaviorists, Holt, suggested that the movements of an individual be defined in terms of "that object, situation, process . . . of which his behavior is a constant function. . . . So in behavior, the flock of birds is not with any accuracy, flying over the green field; it is, more essentially, flying southward" (24, pp. 161–166). This statement, rather paradigmatic of functionalist modes of thought, involves selective description. The preferred hypothesis contains reference to a remote end (south); or, in the words of Heider (21), it sets "distal determination" over "proximal determination," that is, in our case, over description in terms of momentary position (green field).

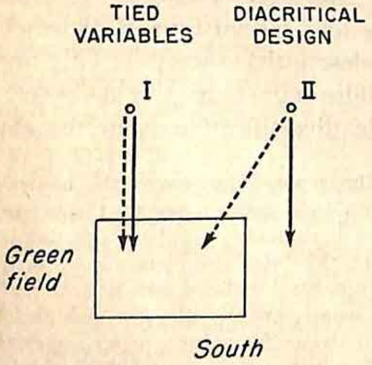


FIG. 1. Systematic designs in the study of behavior constancy (applied to an example from Holt, 24).

In defending his seemingly teleological attitude Holt points to the emphasis some positivist physicists—we may think primarily of Mach—place on the relatively descriptive study of "functional relationships"; these may connect events over space or time regardless of the traditional tracing of causal chains in near-action.

Tied variables. Functionally irrelevant generalizations. Holt's problem of constant function involves generalization. One of the most time-honored traditions in experimentally testing generalizability demands that one, or perhaps a few, conditions be varied in a planful manner decided upon by the experimenter while all others are held constant. The purpose is to assure isolation of the so-called independent variables. For their arbitrary orderliness and confinement such designs may be called "systematic."

For example, we may move the birds backward along the line of their flight, say, to position I in Fig. 1. The birds may persist in their original direction and in this sense show generality of behavior. But it is easily seen that this experiment is irrelevant to Holt's chosen alternative. The two directions, "southward" (solid arrow) and "over the green field" (broken arrow) coincide; allowing for all possible types of response, the two variables involved—south vs. non-south, and green-field vs. non-green-field—are perfectly correlated and thus inseparable so far as the available evidence is concerned. This constitutes artificially induced perfect confounding, and may be labeled "tied-variables" design or, in short, tied design.

Responder replication, that is, repetition with new individuals or with the original individuals at other occasions, is likewise irrelevant to Holt's alternative, regardless of the interindividual or intraindividual consistencies that may be observed. This must be pointed out in view of a rather deeply ingrained trend in psychology of throwing the entire problem of generalization and of statistical significance onto the responder rather than onto the situation (7, pp. 36 ff.; further drastic evidence to this effect, unearthed by Hammond [18, 19], will be referred to in the section on clinical application).

Diacritical confrontation. Splitting of tied clusters vs. isolation of variables. The real testing of Holt's preferred south hypothesis does not even begin until adherence of behavior to the southward direction is made situationally incompatible with adherence to the green field. This could be achieved, say, by moving the birds sideways to position II in Fig. 1. The two arrows issuing from the new starting point are now divergent rather than parallel. The responder is placed at the crossroads and forced to take sides. The previously tied situational factors are now "confronted" (2); we may speak of this variant of systematic design as "diacritical" design.

We soon discover that southwardness is still tied to such factors as the general area of start, temperature and other climatic conditions, topographical landmarks, magnetic cues, and so forth; and so is the greenness of the field to its squareness or size. What we have accomplished in diacritical design is to separate or "split" one original encompassing cluster into two subclusters of tied variables; but we have not really "isolated" our variable as it may have seemed at first glance, and therefore are not yet entitled to speak of its attainment as a constant function.

Perception as constant function. A concrete perceptual problem comparable to Holt's problem of the constancy of southward flight is that of size constancy. The alternative here is between invariance of the response relative to the measured (or computed) sizes of the stimulus impact at the retina (or at a parallel photographic screen) as the proximal variable, which we will call *P*, and invariance relative to the measured sizes of physical bodies underlying this impact as the distal variable, *B*. Size constancy involves the predominant focusing of the response on *B*.

Classical psychophysics as pseudo-univariate design. With distances variant as they are in daily life, there is a certain degree of statistical independence between *B* and *P* despite the existing causal nexus between them. During the classical phase of psychophysics, however—still strongly in evidence today—the implicit design policy was artificially to tie the distal and proximal variables. For size this is achieved by holding the distance from the observer constant. A good example is the Galton bar. The task is molded closely after ordinary physical length measurement except that the lines are laid up lengthwise rather than being superimposed. This creates a tied-variables design which is comparable to case I in Fig. 1. Note that the tying of the two variables is the direct result of a celebrated device of systematic design, the holding constant of a third variable (in our case, distance).

This design may also be cast into the form of a table of presence, or scattergram. Figure 2 presents the major systematic designs in their minimal form, assuming only two levels of strength for each of the situational variables. In the case of the classical tied-variable design the two variables are perfectly correlated. All entries lie along a diagonal. Photographic size (assumed to be plotted vertically as in Fig. 3) is always large when bodily size (plotted horizontally) is large, and vice versa. And judgments correct (or incorrect) concerning one of these variables are

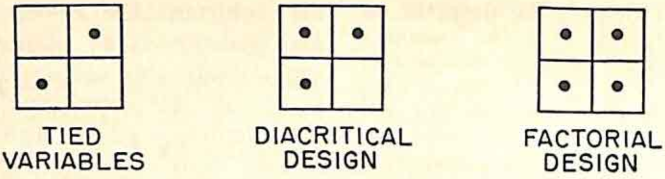


FIG. 2. Minimal scattergrams of systematic designs. Repeatable in k dimensions.

automatically correct (or incorrect) concerning the other, just as Holt's birds, by flying south from position I, automatically must fly over the green field. The classical design therefore precludes decision as to whether the response focuses on proximal (photographic) or distal (bodily) size; nor does it allow the conclusion that the response focuses on either, for that matter.

In the classical phase, the tying of variables was done inadvertently and in considerable naiveté as to the interpretational consequences involved. Since only the bar was directly manipulated and only one scale was read in doing so, the impression prevailed that there was only one independent variable. The notion of univariate design was somewhat obliquely reinforced by the tendency to confine the concept of stimulus to the proximal stimulus variable that had been arranged to vary in unison with the bar. The possibility of varying degrees of dependent variability within the design itself was thus ignored and no provision was made to state the relationships explicitly, so much so that the term "dependent variable" could become synonymous with the organism's eventual responses. Since classical design purports to be univariate yet fails to isolate the distal from the proximal variable, it may be called "pseudo-univariate" design, and in our particular example may be specified as proximo-distally neutral design.

Thing constancy research as a form of multidimensional psychophysics. Factorial design. The diacritical confrontation of distal and proximal size is achieved by removing one side of the bar to a different distance from the observer, say, a smaller one. This has an effect on design analogous to the moving of the birds from position I to position II in Fig. 1. It is now incompatible for the two lengths to be equal bodily and photographically at the same time. The new combination of a relatively small bodily size with a relatively large photographic size injected into the design is plotted in the upper left corner of the center chart in Fig. 2. The forking or parting of the ways which is characteristic of diacritical confrontation may be visualized by assuming the near object to be the Standard and the far one the Comparison, the latter being left to vary along the diagonal.

By projecting the Standard two ways, along a horizontal and along a vertical line, to their respective intersections with the diagonal representing the Comparison series, we obtain two points of objective equality (or POE, formed in analogy to Woodworth's PSE for point of subjective equality [see **7**, **53**]). One represents the variable P and the other B in the experiment. By contrast, in a classical experiment the Standard coincides with a point on the diagonal and thus the two POE's merge in one point, further supporting the erroneous impression of univariate design of which we have spoken above.

The relative allegiance of the response to B or P, or the degree of perceptual "compromise," may be ascertained by inserting added values between (or beyond) the two ideal "poles of intention" (2) along the diagonal (7, p. 17, Fig. 4). The constancy ratio (called Brunswik-ratio by Woodworth) is a simple device to project the obtained PSE's onto the span between the two pole POE's.

Note that the Comparison in psychophysical experiments has often been called *the* "Variable," so that not only classical psychophysics but even our present diacritical experiment may give the superficial impression of unidimensionality, in spite of the fact that the presence of two POE's clearly marks it as a case of multidimensional psychophysics. In this sense we may say that while classical design is pseudo-unidimensional, diacritical design is crypto-multidimensional.

By adding a fourth point, diacritical design becomes the well-known factorial design (14), also shown in Fig. 2. In essence, this is no more than adding a mirror image to diacritical design, with added advantages accruing by virtue of the increased symmetry. Some of the Vienna constancy experiments have employed such a double diacritical design (2, pp. 167 ff.) but they will not be discussed here.

Variate packages and the indefinite regress of systematic design. Diacritical confrontation or factorial design may be carried to several dimensions. Their number is, or at least should be, known to the experimenter; it may be designated by k.

In analogy to what we have said about Holt's example of bird migration, all results remain contingent upon the ties existing within the more or less incidental situational instance from which systematic variation has taken its start. If we use the term "variate" for the specific values along the various variable dimensions, each concrete situation may be regarded as a "variate package." Originally, each of the particulars or variates in the package has equal claim for being singled out in the description, and one or a few factorial separations remove but little of this indeterminacy. Complete "systematic" isolation of one variable as the crucial factor would involve diacritical confrontation with a very large, and in fact indefinite, number of originally tied situational variables.

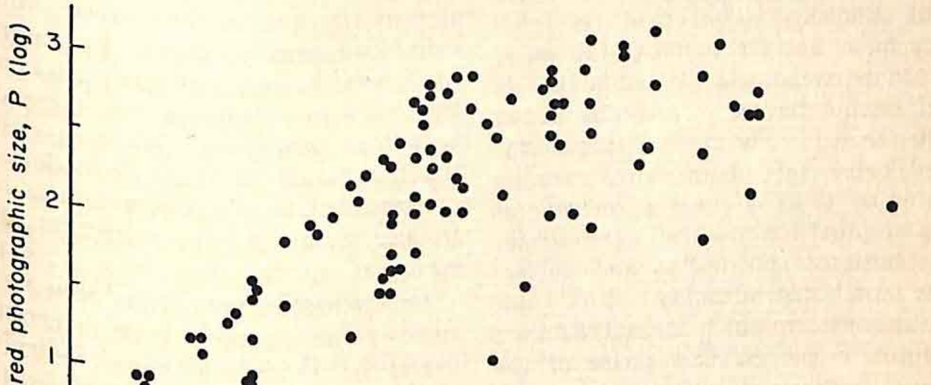


FIG. 3. Scattergram of an example of representative design used in the study of size constancy (adapted from Brunswik, 6, 7). Analogously in x dimensions.

REPRESENTATIVE DESIGN IN THE STUDY OF DISTAL ACHIEVEMENT

From a purely formal point of view the systematic confrontations would have to include the indirect cues for size and distance on which the mechanism of constancy must depend. But it is evident that setting the distal variable against its own instrumentalities would cut out the ground from under the very function whose constancy is to be put to test. Tolman (47) has stressed that all behavior requires the presence of means or "behavior supports." He has also pointed out that the only admissible operational criterion for the testing of "purpose" is the observation of the actual reaching of the end, at least in part of the behavior instances in the class under consideration. But reaching the south or any other distal goal, be it behavioral or perceptual, can obviously become a more or less stabilized function only if the flight of the birds is allowed to take adequate advantage of the natural resources of orientation and locomotion, much as such man-made stabilizers as gun sights must be tested under conditions of practical use.

Ecology and situational instance. Constant psychological function thus is intrinsically limited, or probabilistic, rather than "universal." Flying southward, being right about object sizes, or any other gross or "molar" behavioral or perceptual function can never attain the status of an ironclad and universally applicable so-called "strict" law in the sense in which these laws were idolized in the classical phase of the natural sciences. The basic aim of our initial quotation from Holt requires delimitation of a more specified universe within which the animal is set to operate. This is in line with the "syntactic requiredness" (22) to define all probability in terms of a corresponding reference class or universe. In line with biological usage we will call this universe "ecology." An ecology is defined as the natural-cultural habitat of an individual or group, but is otherwise free of contamination by the system of specific responses. Rather, the ecology is the objective, external potential offered to the organism for survival and its subordinate needs. Nourishment value of foods, as it exists prior to and regardless of its recognition or consumption by the responder, is an example of an ecological variable or set of variables; object size and its system of cues enters via its relevance for manipulation or orientation.

Since the responder merely acts like a catalyst in the definition of the ecology, the ecological environment is not to be confused with Lewin's psychological environment or "life space" which is defined as the reflection of a situation within the response system (see 5).

Situational instances in an ecology are analogous to individuals in a population of responders. Both may be considered as sets of more or less incidental variate packages. The difference is that instances can be taken apart and created at the spur of the moment while individuals usually cannot. But, as we have come to see, a program of functional research demands that they too be left as they come. We must resist the temptation of the systematic experimentalist to interfere, and must introduce a laissez-faire policy for the ecology.

Representative sampling of situations. As we cannot possibly hope to encompass the entire population of individuals in research, but must sample representatively, we must sample instances in the study of functional achievement. Taking the cue from differential psychology, we may transfer the entire formal statistical instrumentarium developed in the study of personality to functional problems as a new content. This will assure, to any desired degree of approximation, a balanced view of psychological function as it comes about by a synopsis of performance under comfortable conditions, manageable vicissitudes, and a due proportion of risks or well-nigh insurmountable odds.

In terms of experimental design there results a combination of constraint and license in which the experimenter is in no more than supervisory control over the adequacy of sampling. There will be a limited range and a characteristic distribution of conditions and condition combinations. If in this manner psychological experiments take on the character of statistical surveys, we may speak of "representative design" (7).

A representative design in perceptual size constancy. Since in representative design the accent is on sampling from an ecology and on the generalizability of functional constancies to this ecology, rather than on sampling and generalizability in reference to a responder population, it was deemed advantageous for demonstration purposes to confine a pilot survey of perceptual size constancy to a single subject. Using n for the number of responders and N for the size of the situation sample we thus had $n = 1$ subject, a graduate student in psychology; there were $N = 93$ object situations (6; 7, pp. 41 ff.). The objects were sampled, in a reasonably random manner, from the sizes that became "figure" to the subject in her daily routine, in package with their natural setting and accompaniment of depth cues.

Ecological validity and dependent variability within the design. Textural ecology. As in all representative design, the design in itself has the character of a result, even though this result concerns the ecology only and is no more than the precondition of psychological investigation proper. The design obtained for our size constancy survey is shown in Fig. 3. The two major distal and proximal stimulus variables are plotted logarithmically, B in terms of millimeters, and P in terms of millimeters at an assumed projection distance of 1 meter.

The manner of covariation, defining what we have called dependent variability within the design, may in first approximation be expressed by a correlation coefficient. In our case the Pearson r between the logarithms of B and P is .70. This coefficient estimates the cue potential of the proximal variable P relative to the distal variable B in the given ecology, in analogy to the way the validity of a test relative to a personality trait may be ascertained for a population of responders. A correlation between ecological variables, one of which is capable of standing in this manner as a probability cue for the other, may thus be labeled "ecological validity." The study of ecological validities, being bivariate correlational, defines what we may call a structural or textural ecology, in contradistinction to the emphasis on unidimensional distribution (of temperature, precipitation, size of population centers, etc.) which is more typical of biological and a part of cultural ecology.

Our particular coefficient indicates that large retinal impacts are somewhat more likely to be caused by relatively large objects, regardless of distance. Considerable as this relationship may seem, it is, as we shall see in a moment, trifling in comparison with the final achievement of the constancy mechanism. This gain becomes possible only by an additional utilization of distance cues. For some of the less valid of the commonly listed perceptual depth criteria, such as vertical position, subdivision of space, and brightness, Seidner has found moderate ecological validities ranging to about .4 (report on preliminary data in 7, pp. 47 ff.). It is easily seen that not even the so-called primary depth cues, such as binocular disparity, are foolproof in our ecology. For example, binocular disparity is present in the stereoscope, yet depth is absent in the underlying reality; in viewing reality through a camera, on the other hand, binocular disparity is absent while depth is present in the chain of causal ancestry.

In our present example the analysis was not carried to an explicit treatment of depth criteria and other context factors. With the use of photographs, such as those Seidner had available for his analysis of depth cues, a great variety of them could be analyzed in a single enterprise. In fact, Fig. 3 must be seen as combining but two out of a practically unknown number, x, variables. Since covariation must be allowed to take its natural course, the different juxtapositions would of course not look exactly alike, as they do in factorial design, but would merely be analogous, forming in the end an x-dimensional space in which all factors could be considered simultaneously.

Quite aside from the avoidance of the pitfalls of systematic design—in which all factors held constant are lost for the investigation, and the resultant tied-variable clusters only lead to confusion—representative design, while cumbersome and laborious, is thus potentially a very economical technique.

Projections of the frequencies in Fig. 3 upon the main axes are shown in Fig. 4. Our sample, restricted to sizes not tilted into the third dimension and thus by-passing the problem of shape constancy, is represented by the solid curves. It is reassuring to find the distribution for *B* fairly normal. (The third graph shows the distribution for distance as related to a dependent distance axis that may be imagined to run across the upper right-hand part of Fig. 3 under 45°, forming a triangle with the main axes; the crowding of both small and intermediate objects along the

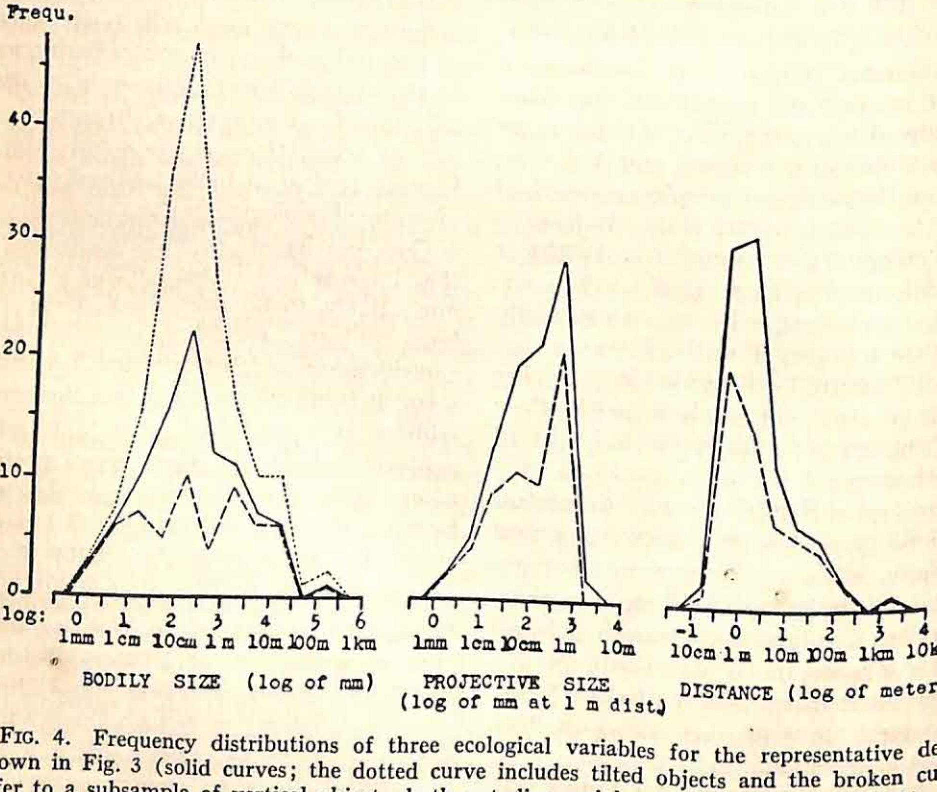


FIG. 4. Frequency distributions of three ecological variables for the representative design shown in Fig. 3 (solid curves; the dotted curve includes tilted objects and the broken curves refer to a subsample of vertical objects, both not discussed here). (From Brunswik, 6.)

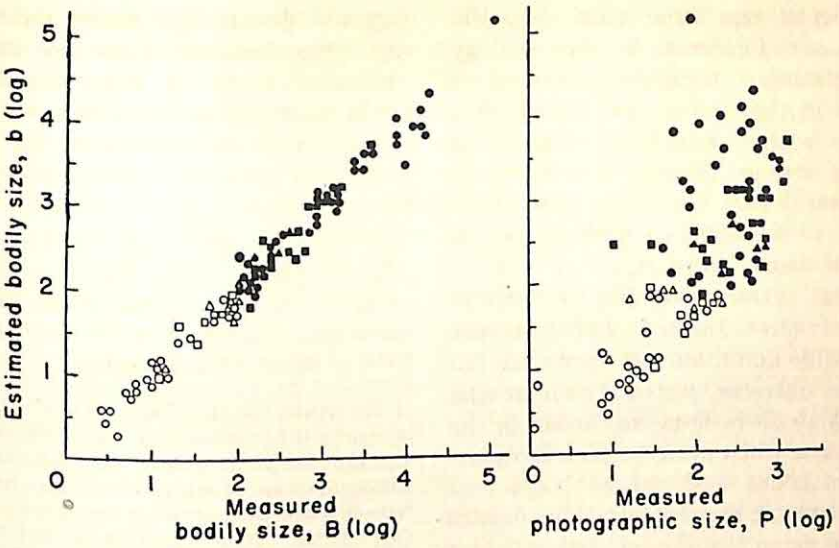


FIG. 5. Scattergrams of perceptual achievement (functional validity) for a subject responding to the representative design shown in Fig. 3. (Open symbols are used here to designate objects smaller than 10 cm., see below; the differences in shape are irrelevant in the present context.) (Adapted from Brunswik, 6, 7.)

10-inch limit of near vision leads to skewed distribution and is the chief source of the curvilinearity that may be observed in Fig. 3.)

Functional validity (achievement). We now turn to covariation between the distal stimulus and the response variable which will define perception in its approximation to "constant function" as understood by Holt. Figure 5, left, shows the stimulus-response scattergram for the logarithms of the perceptual estimates, *b* (so labeled because they were given in the natural, naive-realistic attitude toward bodily size, *B*) as plotted against the logarithms of *B*. The entries cling fairly close to the diagonal, much more closely than the major cue variable, *P*, was found to cling to *B* in Fig. 3. Appraising the maximum errors committed in the particular estimates, however, we find that some entries deviate as much as one-half of a power of 10 from the diagonal, corresponding to about three- to fourfold over- or underestimation of the length in question.

A generic summary description of the degree of perfection of size constancy may again be sought in the correlation coefficient. This represents what may be called the "functional validity" (7) or "achievement" (*Leistung*, 2; for definition see 5, p. 255) of the class of responses, *b*, relative to the distal stimulus variable, *B*. For the total sample of 93 situations this correlation is close to .99 (more accurately, .987 when computed from ungrouped logarithms with three significant places [see 7, p. 44]), in spite of the occasional sizable errors referred to above.

By contrast, the functional validity of *b* relative to photographic size, *P* (Fig. 5, right), is only .73; in addition, this is quite close to the purely ecological association of *B* and *P*, .70, so that little independent focusing of *b* on *P* is indicated. Perceptual restructuring toward photographic size (painter's attitude, *p*) raises the correlation of the response with *P* no more than to .85. Other attitudes, deliberately inviting intellectually controlled judgment ("betting") and thus the "stimulus error," will not be discussed here, nor shall we go into the question as to whether or not the estimates *b* and *p* are purely perceptual.

Ecological replication and ecological significance of differences. The analogy to the statistics of individual differences may be carried still further. We may consider each new chance selection of items like those in Fig. 3 as an "ecological replication" of the original sample. In representative design the traditional quest for repetition under "identical" external conditions ceases to be legitimate. Instead, we must seek comparable conditions as drawn from a common universe, just as in differential psychology there is no repetition in the strict sense when other or even the same individuals are used.

Problems of the ecological significance of differences in the statistical sense may be raised, and handled, accordingly. The size of the sample to be used in computation is now given by the number of situations, N, rather than by the number of subjects, n. In this sense the ecological generalizability of distal rather than proximal focusing, in the area of size constancy with distance variant, has been established for our one subject; and even the more crucial in a wide array of relatively moderate ecological validity coefficients of depth cues, referred to above from the material of Seidner, have been found significantly above zero.

In the strict, technical sense, our representative survey of size constancy is completely void of interindividual generalizability. An approach to responder generality was made by using the recorder as a second subject; his distal functional validity was .993, quite close to the .987 for our subject. Furthermore, Dukes (12) has independently obtained a correlation of .991 for a six-year-old boy, using a different sample ($N = 67$) and a somewhat different technique.

It may well be that in many contexts individuals in a population are more homogeneous or stereotyped than are situations in an ecology, and that the ascertainment of ecological generality may be a more challenging task than that of responder-populational generality. Harvey needed only one person to demonstrate the circulation of the blood, and Ebbinghaus needed only himself as a subject to lay the foundation for much of modern learning theory, ecologically narrow as this theory may be.

To complete our analogies between responder-populational and ecological generalization problems, we may set physical size measurement and intuitive or critical estimates of size in analogy to a battery of tests given, not to n persons, but to N situations of the environment. Each type of observation or attitude represents one of these tests. The problem of the degree of "objectivity" of various classes of observation, and thus of certain scientific approaches, can then be handled statistically in terms of their inter- or intraindividual "observational reliability" as tests, in which a sample of situations has taken the place of a sample of persons (see 7, p. 33 and 8, pp. 11 f.).

Representative separation and mathematical isolation of variables. It will be remembered that under systematic design a true isolation of variables can be achieved only by a virtually infinite series of diacritical confrontations or factorial designs piled upon one another. Representative design, while not laying claim to full isolation, separates variables to the extent to which they are separated in the particular ecology but no further, and does not tolerate any artificial perfect tyings (or untyings) between variables. Variables may thus be said to be "representatively separated." From a systematic point of view, a good deal of spuriousness remains built into the textures studied under this policy, samples as these textures are of an ecology that likewise is complexly textured. Representative design is not afraid of this spuriousness; in fact, it welcomes it for the sake of the behavior supports it allows in the execution of the functional approach.

The challenge of further isolation must be met by after-the-fact, mathematical means, as in the study of individual differences. For example, we may use partial correlation as a mathematical means of holding constant a certain variable. Partialing out *P* from our above correlation between *b* and *B* (and thus in effect reducing the ecological validity of *P* from .70 to 0) still yields a functional validity as high as .98; whereas factually eliminating the sizes under 10 cm. in Fig. 3—to the right of log $B = 2$—and thus reducing the ecological validity of *P* from .70 to .14 in a quasi-systematic move, reduces the functional validity to .95 (see also 7, Fig. 9 and the accompanying text).

It must also be noted that, in contradistinction to systematic design, the process of analysis may be stopped at any point, falling back on the nonreductive aim of functional research, together with the assurance that the unresolved part of the associations is safely within the fold of the ecology to which the investigation has been geared from the beginning.

SYSTEMATIC EXPERIMENTS WITH REPRESENTATIVE FEATURES

We now turn to certain experimental policies, some of them common, which may be considered transitory between systematic and representative design. Representative features may be injected in otherwise systematic designs in a variety of ways.

Quasi-representative choice of variables and of their variation or covariation. Some measure of representativeness may be achieved by the choice of variables with particular life relevance, such as "value," as a factor in constant function. Since ecology embraces cultural norms held valid by the law enforcement policies of a society, along with those connected with physical law or geographical contingency, monetary value becomes a challenge to perceptual attainment on a par with other object properties. Cognitive considerations of this kind, rather than the emotive or motivational aspects of value, per se, prompted the study of what could be called perceptual value constancy. In experiments concerning the apparent numerosity of stamps and of coins, Zuk-Kardos and Fazil (reported in 2, pp. 140–150) as well as Ansbacher (referred to in 7) found number constancy with value variant and value constancy with number variant fairly high, although tainted with compromise between the two variables.

Another quasi-representative step is to gear the manner of variation or of covariation between variables to the general scheme of natural conditions in a planfully controlled way, as when the association between a certain cue or means and the object or reward is made probabilistic rather than absolute (4, 10).

Experiments centered about an exemplary instance. Successive omission vs. successive accumulation of cues. A certain effort toward representativeness is discernible whenever a "lifelike" situation is taken as the starting point of the experiment. In the field of size perception such experiments are likely to abandon, at least in some of their phases, the chin rests, darkrooms, screens with small openings, alleys of edges without thickness, or other laboratory paraphernalia in vogue during the late nineteenth century. This liberalization owes much to David Katz (29) and his work in another area of perceptual constancy—color constancy with illumination variant. For size constancy, a study by Holaday (23; see also 7, p. 23), and the studies by Holway and Boring (25), by Gibson (16), and to some extent one by Joynson (28) have proceeded by essentially the same functional scheme.

In each case there is what we may call a core or "exemplar" situation—somewhat arbitrarily chosen, to be sure

—which contains a fairly natural array of cues. Superimposed upon this core is a greater or lesser number of systematic variations so that the design is still fundamentally systematic in its ramifications. The systematic part usually effects what may be called "successive omission" from the originally unknown, or at least not fully scrutinized, array of cues, while the more nearly classical approach tends to eliminate all natural cues, then building up from nothing in a technique of "successive accumulation" (7, pp. 22 ff.).

By and large the functional studies of size constancy have borne out Hering's old assumption of "approximate size-constancy" according to which more distant objects tend to be slightly underestimated (although not nearly as much as would correspond to the shrinkage of their photographic size). But it is significant, from the methodological point of view, that in some of the studies the most favorable cue conditions yielded on the average slight over- rather than undercompensation for distance ("overconstancy"), and that the over-all level of "compromise" varies considerably from study to study. In addition, laboratory experiments inadvertently employing certain atypical contexts or backgrounds will yield drastically different results, not to speak of essentially traditional experiments that programmatically employ grossly distortive configurations, like the situations recently created by Ames (27).

Under the aspect of representative design, all these systematic experiments must be viewed as ecological single cases or "instances" with artificial elaboration that leaves a large portion of the core elements untouched. Each of these experiments is indisputable in its results, but at the same time is of unscrutinized ecological generalizability. The variate packages constituting these experiments or experimental settings may be projected somewhere in the x-dimensional manifold of representative design which we have exemplified for two dimensions in Fig. 3. Each imaginary point, or small, orderly group of points, in such a space represents a potential systematic experiment. Mostly there is little technical basis for telling whether a given experiment is an ecological normal, located in the midst of a crowd of natural instances, or whether it is more like a bearded lady at the fringes of reality, or perhaps like a mere homunculus of the laboratory out in the blank.

As a matter of principle, individual sample situations, no matter how lifelike, cannot answer the funtional problem as to the degree of perceptual constancy, even though by the use of responder replication or by systematic variation their results may become generalizable in certain directions and standardized for testing purposes. Only representative design can answer this problem. By a set of analyses not reflected in the figures of the present paper, the ecological generalizability of the principle of perceptual compromise, or of "approximate" size constancy as originally suggested by Hering, has been established for our subject along with the broader principle of distal rather than proximal focusing.

Canvassing as accidental quota sampling of the ecology. In still other cases an entire array of individual systematic experiments may appear to be laid out after quasi-representative principles, so as to cover the ecology or the mediational pathways by a vaguely conceived "one of a kind" rule. This tacit sampling procedure forms a counterpart to what polling statisticians might describe as a most rudimentary form of stratified, "quota" or proportionate sampling and is usually of the highly erratic type sometimes labeled "accidental" in statistics. We will speak of this primitive type of coverage of the ecology as "canvassing."

In the field of the perceptual constancies, the attempted extension of basic principles, such as that of distal focusing or of compromise, from size to shape to color constancy, is an example of canvassing. A cross-departmental extension to "loudness constancy" with distance of sound source variant was undertaken by Mohrmann (37), yielding generally similar results. Kinesthetic experiments with falling bodies, with and without the aid of visual cues, have established the functioning of a perceptual "weight constancy," with speed (and thus with kinetic energy) variant (2, pp. 161 f.). The above-cited experiments on value constancy further augment the picture of canvassing of the perceptual constancies.

Truncated factorial design and an impasse in the analysis of variance. The basic intent of representative design is toward design proper; that is, it concerns the employment of a statistical selection device for the stimulus sample; only secondarily is it concerned with statistics as a procedure of evaluation. Its relevance to analysis of variance is thus indirect, and derives solely from the fact that this evaluation technique is an adjunct to factorial design, conceived in its shadow and thus wrought with its inherent inadequacies.

At first glance factorial design may appear ideal, since all plots are filled, while in representative design normally only part of the plots is filled; representative designs would thus seem readily extractable from factorial designs if anyone should wish to extract them, but not vice versa. The first catch is, however, that some of the intercombinations of variates may be incompatible in nature or otherwise grossly unrealistic.

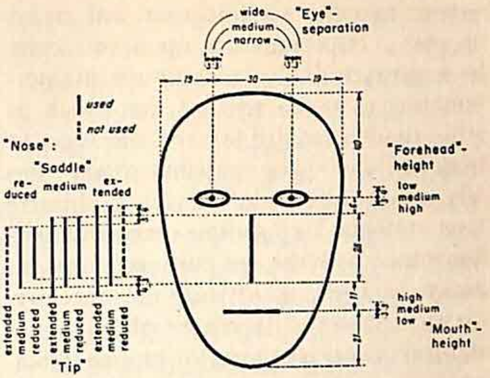


FIG. 6. Five-factorial schematized face (adapted from Brunswik and Reiter, 11).

An example is furnished by a study of the perceptual impression values of schematized faces by Brunswik and Reiter (2, 11; for a summary and discussion see 7, pp. 40 f.). In effect, this was an early factorial design, with a $3 \times 3 \times 3 \times 3 \times 3$ layout (Fig. 6). Fortunately for our argument, we were naive enough to depart from the complete array of intercombinations so far as two of the five "facial" variables chosen are concerned. The inadvertently unorthodox part is the truncated treatment of the "nose," as plotted in Fig. 7. Two opposite corners are left empty so that there are only seven noses instead of nine, making for a total of $3 \times 3 \times 3 \times 7 = 189$ facial schemata. The result is an oblique relationship in the artificial ecology of the design which is not unlike the representative ecological correlation shown in Fig. 3 for size constancy, even though with a strange orderly tint that bespeaks the systematic origin of the stimulus distribution.

The reason for the departure from strict orderliness of the design was

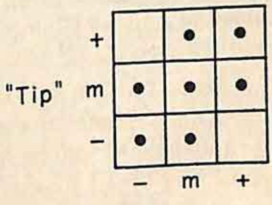


FIG. 7. Truncated treatment of two out of the five variables shown in Fig. 6.

given by the desire—still somewhat vague at the time—to achieve better representativeness. One of the desiderata was to make the schemata look as "facelike" as feasible; the omission of two of the nine possible noses was prompted by their unusually bizarre and ridiculous appearance which threatened to spoil the seriousness of an already precarious attitude on the part of the subjects. In many other psychological contexts certain intercombinations may be unrealistic or disruptive in similar or in some other ways.

Like the design, our original evaluation was makeshift. Years later the main effects and some of the first-order interactions were found significant in terms of analysis of variance as applied to the 189 original impressions of intelligence (composite from 10 adult subjects), although this of course implies little concerning the ecological generalizability to live faces. For our schemata it appeared that impressions of intelligence are aided primarily by high or medium forehead, medium mouth, and medium nose.

In the analysis of the impression value of the "nose" the omission of the two cells necessitated—as my statistician colleagues, Drs. Rheem F. Jarrett and Robert Rollin, informed me—the employment of such *ad hoc* procedures as the "missing plot" or the "unequal numbers in cells" technique (30, pp. 220 ff.). These techniques are not only artificial and rather uncommon; what is more, they involve the somewhat hypocritical pretense that part of the data was lost, or at least that the material is not quite as it ought to be. Under representative design, involving as it does variate packages in many more dimensions than factorial design could handle practically, oblique distributions are not only legitimate but will in most cases be dictated by nature or by culture, and thus be mandatory. Evaluation techniques will have to be fitted to the materials that may be obtained under such design, rather than vice versa.

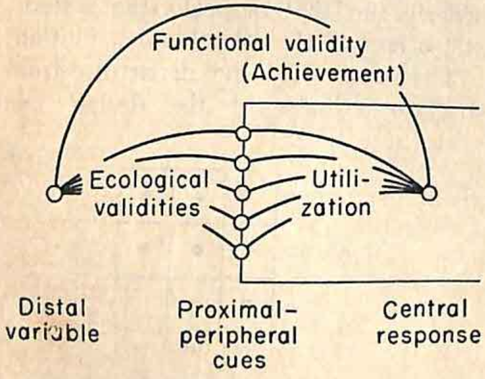


FIG. 8. The lens model as applied to perceptual constancy (adapted from Brunswik, 8).

Major Varieties of Psychological Theorizing

We now turn to the implications of our above considerations for psychological theory. Three types of psychological theorizing will be discussed: (*a*) theory as the ratiomorphic explication of probabilistic functional—notably distal—achievement and of its strategy; (*b*) theory as the customary nomothetic and more or less molecular reduction of function; and (*c*) theory as comparative methodology, both within psychology and between psychology and the other sciences.

Theory as a ratiomorphic model of functional achievement and its strategy. Focusing and vicarious mediation. The representative study of distal constant function, as reported above, is schematically described in Fig. 8. The wide-arching functional validity coefficient constitutes a generalized statement of the organism's perfection in the attainment of a given distal variable; it falls under the concept of "descriptive theory" by virtue of its generality alone. In addition, an inventory or mapping of the full array of outpost variables attained by an individual or species, as well as of the degree of such focusing ("psychology in terms of objects," 2, 3), promises to lead to further generalizations concerning preferential focusing on distal vs. proximal variables at large, concerning the relationship between life relevance and distality, and concerning related biological-functional problems.

As is further indicated in Fig. 8, the over-all functional arc may be broken down into an extrasystemic and an intrasystemic constituent, called respectively, ecological validity and utilization. The general pattern of the mediational strategy of the organism is predicated upon the limited ecological validity or trustworthiness of cues which we have observed earlier in this paper. This forces a probabilistic strategy upon the organism. To improve its bet, it must accumulate and combine cues. Thus we arrive at a better understanding of the principle of mutual substitutability, or "vicarious functioning," of means (or cues) for each other which Hunter, Tolman (47, ch. i), and most other behaviorists looking for a structural criterion have incorporated in their basic definitions of behavior or purpose (see 8, ch. ii). Hence the lens-like model in Fig. 8, which may be taken to represent the basic unit of psychological functioning. No matter how much the attainment is improved, however, distal function remains inherently probabilistic.

In the light of this model all "constant," or rather, quasi-constant function, be it "intuitive" or explicit, can be explicated as a statistical reasoning process remindful of Helmholtz's "unconscious inference," albeit without its introspectionistic and perfectionistic overtones. Forming a Latin-Greek hybrid, we may speak of this as a "ratiomorphic" theory of achievement (9).

One of the most important aspects of such a theory concerns the relationship between ecological validity and utilization. Ideally, cues should be utilized in accordance with their validity. But here we must inject, among other things to be brought up later in this paper, the element of "cost" to the organism, just as we must ask for the cost of an automobile along with its efficiency in budgeting our expenditures. Functional theory here takes on certain features of economic theory.

Theory as nomothetic reduction. We have left the discussion of the systematic approach, in favor of that of representative design, at the point where the former was about to embark on a diacritical confrontation of constant function with its own cues or other mediating instrumentalities. This type of confrontation has been practiced extensively in psychology, notably in its more classical phases. In the founding period of behaviorism, Watson and Lashley (50) set out to transport homing birds in closed cages over distances sometimes far outside their natural range. Others have since used anesthetics, faradic cages, or rotation on a phonograph turntable in the dark throughout the journey (45). The artificially distortive laboratory experiments on the perception of size and distance to which we have referred above as preceding, and in part paralleling, the quasi-representative and representative studies of the last quarter century fall in the same category.

The objective of this type of behavior or perception study is to trace performance step by step to identifiable processes, cues, or tracks of mediation; in the end they are to be "reduced" to the "laws" of one of the more microscopic, more "fundamental" disciplines, notably physiology. Traditionally, many psychologists have seen in such a reductive explanation of behavior the major task of psychological theory. The nomothetic behaviorism of Hull and of his sympathizers, mentioned at the beginning of this paper, reveals its reductionist aim most clearly in the use of a "physiologizing" terminology.

To the study of distal function and of its grand strategy the reductive approach adds the study of tactics; to the study of achievement and of its macromediation, both of which fall within the province of functional-representative design, it adds the study of micromediation, which falls in the province of nomothetic-systematic design. The functional approach has its place mainly in the appraisal of the interplay and relative contribution or weight of factors in the adjustment to a given ecology, while the reductive approach reveals the technological details of the machinery that brings about such adjustment.

Theory as comparative methodology: observational unity and thematic diversity of science. The injection of physiology into the discussion brings us to a branch of psychological theory which overlaps with the philosophy of science and is best labeled comparative methodology of science. One of the major concerns of such a science of science, or metatheory, is with the basic unity of the sciences. In present-day psychological discussion this problem is subsumed mainly under the watchword "operationism." Here it is often forgotten that the basic requirement for scientific exactitude is a relatively modest one, and in a certain sense a more commonplace one than anticipated. It involves no more than the inter- and intrasubjective univocality of observation and communication which is sometimes called "methodological physicalism" (8), but should better be specified as observational or procedural physicalism; but it does not include the univocality of prediction which is the major *raison d'être* of the nomothetic approach (5).

We may therefore take the position that in the end the unity of science is better served by allowing the reaffirmation or elaboration of this unity to be superseded by a working out of the thematic diversity of the sciences within the minimum common platform. This diversity of themata involves both the aim of the different disciplines and the designs capable of serving these aims.

Thematic physicalism and the nomothetic-reductionist-systematic syndrome. The different explicit and implicit trends toward the unification of science which have dominated the last half century have been under the spell of a somewhat stereotyped image of physics. The thematic element in this cliché may be explicated as the emphasis on the "general," and notably on the strict, univocal regularities or laws which possess universal applicability; this is also known as the nomothetic approach. Universality of law presupposes homogeneity of the universe; hence it matters little where and when and over how large an area a phenomenon is studied. Experimental design may thus safely be left to the convenience and liking of the experimenter and thus become systematic. And, as Mises (36) has pointed out in discussing probabilism in physics, macrolaws have their origin in differential equations, that is, in principles conceived for minute space-time splinters. The triad of nomothetic aim, microreductionist procedure, and systematic design which we have come to recognize as a syndrome in traditional psychology thus is revealed as an emulation of a pattern indigenous to the specific thema of physics.

As has been pointed out in greater detail elsewhere (8, sec. 9), the basically elementistic character of physical law is not obviated by, and has little to do with, the empirical fact of a probabilistic microstructure in thermodynamics or within the atom. Nor must such structures be seen as related to, or in any way supporting, our arguments for the probability approach in psychology; functional psychology is macroprobabilistic in that the identity of the individual case is maintained (as in a scattergram) while it is lost in the physical macrolaw which for ordinary practical purposes is absolute.

General vs. particular. Physical law, geographic fact, ecological correlation. The first to warn against the overestimation of the general over the particular in science was Windelband (52). In a somewhat stilted application to personality and to the humanities, he was led to suggest the well-known distinction between nomothetic and idiographic disciplines. Unfortunately, the latter term encompasses both the low-brow, strictly enumerative approach to historic-geographic fact seen in isolation, and the extremely high-brow emphasis on the "unique" lawfulness of the individual or culture; more properly, the latter case should therefore be labelled the "idionomothetic" approach.

Within the natural sciences, an example of the purely enumerative approach is given by those branches of geography that deal with topographical mapping. The fact that no application of the general physical laws is possible without the constants, parameters and boundary conditions furnished by geographic types of information is frequently neglected in theoretical discussion. Except at the level of control ordinarily accessible only to the physicist observer, these constants are not available (at least not fully so) to the responding organism. Hence the chains from select distal to proximal to central variables in perception are chains of (probable) partial, rather than of total, causation. The universal lawfulness of the world is of limited comfort to the perceiver or behaver not in a position to apply these laws, and he therefore must rely largely on whatever snitches of particular or semigeneralized information he may be able to assemble. This is what we meant earlier in this paper by the assertion that ordinarily organisms must behave as if in a semierratic ecology.

With data from our representative size-constancy survey described above, the relation between physical law and ecological correlation is illustrated in Table 1. The first row of ecological validities has been introduced above in discussing Fig. 3, and the remaining two pairs are added here from our material. Partial correlations derived from either of the two columns of three co-

TABLE 1

TEXTURAL ECOLOGY AND PHYSICS

(Adapted from Brunswik 7, Fig. 10)

Variables Correlated	Ecological Validities		Nomothetic Approach		
			Partial Correlation		Law of Physical Optics
	Full Sample ($N = 93$)	Sizes over 10 cm. ($N = 59$)	Variable "Held Constant"	Coefficient Obtained	
$B \times P$	.70	.14	D	1.00	$B = PD$
$B \times D$	.77	.88	P	1.00	
$D \times P$	.08	−.34	B	−1.00	

efficients yield (in as close approximation as may be expected) the perfect positive or negative correlations listed in the third column of figures. These in turn reflect the well-known proportionality law, given still further to the right. Since derivation is possible from all alternate ecologies (as exemplified by our sample of 93 vs. the subsample of 59 situations mentioned earlier), it is evident that textural ecology adds valuable probabilistic information to the vastly distilled relational information incorporated in the universal laws of physics. Essentially this is information abstracted and summarized from the geographies that make up the ecology, and is information of the type finite organisms may best be able to absorb in learning, notably in probability discrimination with partial reinforcement.

It is for the same reasons that the laws of triangulation, which underlie the binocular depth mechanism and which only recently have once more been the starting point for nomothetic treatment of this mechanism (32; see also 17), are of somewhat academic interest so far as the actual perceiver is concerned. And for similar reasons we stressed, earlier in this paper, the limited validity of this mechanism within our cultural ecology, in which the perceiver is routinely exposed to optical instruments as well as to flat pictures as substitute means of access to three-dimensional reality.

Nonfunctionalistic (*intraorganismic*) *uses of probability in psychology.* It will be remembered that in our molar-functional view of organism-environment interplay, uncertainty is a feature of the relationships between the organism and the distal environment, to wit, of proximal-distal relationships in the case of ecological validity and of central-distal relationships in the case of functional validity; uncertainty is not seen as a necessary feature of intraorganismic processes. In a certain sense this is perhaps in reversal of a tradition which has seen nothing but law in the environment, but was indeterministic or vitalistic so far as the reacting organism is concerned.

Some recent movements have viewed processes within the organism from the probability point of view. While probabilistic functionalism is inherently intersystemic, their approach is intrasystemic. We have no quarrel with them except that we must make it clear that they are confined to problem conceptions which are oversimplified from the standpoint of molar functional psychology, encapsulated as they are within the boundaries of the organism.

Intraregional statistical approaches in psychology may be classified under three major headings: central, peripheral, and peripheral-central. An example of a central statistical theory is the study of "random nervous nets" as developed in the framework of mathematical biophysics by McCulloch and Pitts (33), later in collaboration with Landahl (31). Concern is not with "this" neurone synapsing on "that" one, but rather with "gross" distribution of tendencies and probabilities associated with points or regions in the net. This statistical "rather than deterministic" biophysics was used by Rashevsky (39) as an underpinning for his fundamental equations of mathematical biophysics much in the manner in which the kinetic theory of gases serves as a reductive support to thermodynamics. Rapoport and Shimbel (38, 42) have extended the theory to the dynamics of social interaction.

An application of probability to a peripheral problem is given by the "statistical behavioristics" of Miller and Frick (35). The exposition is encapsulated within the single region of overt responses. The study of "stochastic" (i.e., to a certain extent predictable) word sequences in common English which has come in vogue recently is more or less bodily transferred to an analysis of human "courses of action" —which are related to the "strategies" in the playing of games (49)—and of the "dependent probabilities" resulting from the fact that the preceding occurrence of a response does not always return the system to the original state. Here the only cue considered is given by the preceding time series of events of the same kind. This procedure is not indigenous nor particularly congenial to the thematic content of psychology. Rather, it is somewhat mechanically transplanted from those segments of cybernetics in which unidimensional sequential distribution and correlation are predominant (51)—another instance of falling for ready-made gadgets, even though at a more elevated level.

Skinner's (43) concept of "probability of response" may likewise be classified under the heading of peripheral encapsulation, the term being no more than a fanciful expression for (relative) frequency of response as one of the traditional scores in learning experiments involving the motor output. The "statistical learning theory" of Estes (13) is an example of a peripheral-central theory, or perhaps of a peripheral-peripheral theory crossing through the entire organism from input to output. Based as it is on evidence from systematic experiments, there is reason to doubt that the "behavior samples" and "statistical samples of environmental events" which are linked by the theory are even tacitly envisaged as representative samples.

Rudimentary emergence of the concept of cue in cybernetics and the theory of telecommunication. Outside of psychology, and within movements which in some of their more obvious aspects have become templates for most of the intraorganismic probabilism just cited, a first step in the direction of the bivariate type of correlation analysis which is covered by the concept of ecological validity is made by Wiener (51) in his analysis of double time series. Such series are, as Wiener points out, most conspicuous in economico-sociological and meteorologico-geophysical applications, since in both instances the relative lead of one time series with respect to another may well give much more information concerning the past of the second than of its own. For example, on account of the general eastward movement of the weather, Chicago weather may be more important in the forecasting of Boston weather than Boston weather itself. It will be noted that even here the comparison stays within the same kind or physical denomination ("weather"); in such cases comparisons or correlations are not yet genuinely bivariate as are those between proximal cues and objects, or between means and goal attainments.

A more direct exposition of those mathematical principles of communication which are of particular relevance to the understanding of focusing by vicarious functioning as it occurs in psychological mechanisms has been given by Shannon and Weaver (41). In terms of the vocabulary of the special brand of telecommunication engineering involving semicontrolled media, to which the theory has been geared, perceptual cues and behavioral means are like "signals" in "coded messages." The mediating channels are contaminated with interferences or constraints of their own. The result is equivocation. It is then "not in general possible to reconstruct the message with certainty by any operation on the signal." Shannon's diagram showing the fanning out of "reasonable causes" (messages, inputs) for a given "high probability received signal" or effect, and of "reasonable effects" (signals, outputs) from a given "high probability message" or cause in a channel, bears formal resemblance to the equivocal types of coupling between intra- and extraorganismic regions to which this writer called attention twenty years ago (2, Fig. 2) and which can also be read into our diagram of the lens model in Fig. 8.

Whenever the "capacity" of a channel is less than the richness of variability of the source from which it accepts messages, the channel is "overloaded." Then it is

> impossible to devise codes which reduce the error frequency as low as one may please. . . . However clever one is with the coding process, it will always be true that after the signal is received there remains some undesirable (noise) uncertainty about what the message was (41, p. 111).

We may add that, in quite the same manner, the crux of organismic adjustment obviously lies in the fact that distal perceptual and behavioral mediation must, by the nature of things, in the general case rely on overloaded channels, with the ensuing limited dependability of all achievement mechanisms. And we may note that at least part of the trouble lies with the overloading and noise in the external rather than the internal medium.

Shannon and Weaver point to one means by which the chances of error can be decreased, however. This is "redundancy," as exemplified by, but by no means restricted to, verbal repetitiveness. When there is noise there is some real advantage in not using a coding process that eliminates all the redundancy, for the remaining redundancy helps combat the uncertainty of transmission. The reader will recognize that the vicarious functioning of cues and of means—functional patterns we came to acknowledge as the backbone of stabilized achievement—may be viewed as special cases of receiving or sending messages through redundant, even though not literally repetitive, channels. The probability of error, given by the variety of possible causes or effects that could result in, or be produced by, the type of event in question, can thus be minimized; this is the case in the gain of the over-all functional validity (.99) over the ecological validity of the major retinal cue (.70) in our representative survey of size constancy referred to above, in which the organism acts as an intuitive statistician.[3]

A suggestion of extending the theory of communication to multivariate patterns of mediation has recently been made by McGill (34). It is hoped that this will open up the full scope of vicarious functioning to formal treatment.

CLINICAL AND RELATED ASPECTS OF REPRESENTATIVE DESIGN

There are several ways in which representative design has become involved with social and clinical psychology. Two of them deal with sampling aspects exclusively, and another two concern both design and the functional theory of the vicarious functioning of cues.

---

[3] The present statistical application of redundancy is not to be confused with one recently suggested by Attneave (1). In his case the concern is with the exploitation (extrapolation) of *strict* law as it holds over limited stretches of space (or time), say, along part of the contour of an ink bottle. In an earlier context, this writer has spoken of such regularities of limited scope as "local laws" (*Lokalgesetze*, 2, pp. 209, 212). By virtue of their strictness their treatment falls under our above-mentioned heading of an idionomothetic approach; its only relation to the probabilistic approach lies in the fact that the area to which a local law is limited may be considered as a subecology, and that the organism may have to proceed by basically similar mechanisms to ascertain either of the two types of limited regularities, the strict or the probabilistic.

Representative sampling of persons in the rôle of objects. The case in which not only the responding subjects but also the stimulus objects are persons furnishes perhaps the most obvious demonstration of the necessity for representative design. It is for this symmetry of subject and object that such media as the social perception of personality from photographs seem to offer the best chance of convincing the designer that there must be a sizable N of social objects alongside the customary sizable n of subjects (judges). Yet a survey of the respective literature revealed that, apparently by force of a somewhat thoughtless, content-bound tradition, the object N is on the whole pitifully inadequate in comparison with the subject n, with fallacies of generalization ensuing (7, ch. vi and p. 38).

Hammond has been especially astute in exposing comparable lapses in the logic of design and of evaluation in other areas of social psychology (18) and in clinical psychology, notably in the study of the effect of the sex or personality of the examiner upon Rorschach and other projective test results (19)—even on the part of standard texts in psychological statistics. It turned out that it was a surprisingly widespread practice to apply only one (responder-populational) test of significance using n, and to tacitly consider this test to cover ecological generality in the same breath.

Representative sampling of stimulus configurations as test situations. Concerning representativeness as to the purely physical stimulus configurations used in testing, test designers have so far not pressed beyond the stage designated as canvassing, earlier in this paper. At best, tests were either picked to reflect roughly the current distribution of systematic laboratory experimentation—in particularly fortunate cases, such as in Thurstone's factorial study of perception (46), with an emphasis on more recent and more complex types of experimental problems—or there was an informal effort to assemble a battery of "close-to-life" situations. If the tests were selected from a pool of existing tests the term "sampling" was occasionally applied, but it was usually forgotten that sampling from an artifact, even when this sampling itself should be random, merely perpetuates earlier bias (7, pp. 50 f.).

Similar considerations apply to the customary so-called sampling of behavior, of acts, or of traits from a trait universe (including the construction of adjective check lists), so long as the respective universes or sampling procedures are but informally scrutinized.[4] The degree of representativeness is in each of these cases determined not by the most but by the least representative step in the chain of defining the universe and of drawing the sample. Since trait sampling concerns the responder rather than the ecology, and hence is not part of representative design proper, this is added here merely parenthetically.

From a technical point of view it is not sufficient to have one close-to-life test, or even many of them. As we have pointed out in the main part of this paper, there must be balanced coverage of life. A mere multitude of situations may be designated as M tests, reserving N for the size of truly representative ecological samples. (M is chosen as the letter preceding N, or as the first letter in *many;* the capital letter is used in slight departure from the practice suggested earlier [7, pp. 34 f.]). Only for a true N can technical tests of ecological significance as envisaged in representative design be applied.

---

[4] Guttman states, seemingly without being much concerned about the underlying problem, that "the processes of sampling people and of sampling items are not at all identical; random sampling, stratified or not, is used for the first, but is not applicable to item construction" (44, p. 54).

Such stimulus configurations as the Rorschach inkblots are biased variate packages, in the sense of the word as defined earlier in this paper. They invite what Sander (40) has called "realistically-meaningful" (*sinnhaft-bedeutungsvoll*) or "ontotropic" tendencies, in contradistinction to the "geometric-ornamental" or "eidotropic" tendencies (toward *Prägnanz* or "good form," such as abstract circles, squares, or other purely geometric formations) as described by the gestalt psychologists proper. Since these ontotropic tendencies are considered to represent a higher developmental stage than the formalistic eidotropic tendencies, both in ontogenesis and in the "actualgenesis" of the instantaneous perceptual experience, a more genuine sampling of stimulus patterns and the setting up of more representative evaluation categories may become diagnostically valuable.

Allowance of vicarious functioning as a representative feature in test construction. As we have seen, one of the most important principles of functional theory is that of vicarious functioning. Its earliest recognition stems not from academic psychology but from one of the theoretical antecedents of clinical psychology, psychoanalysis. Under the direct influence of psychoanalysis, Frenkel-Brunswik (15) has developed the study of "alternative manifestations" of the latent dynamic structure of needs on a statistical basis, using multiple correlation with overt behavior elements; the way to the rational reconstruction of clinical intuition which is asked for in another line of Vienna tradition, logical empiricism, was thus opened. The relationships that may or may not exist with Lazarsfeld's subsequent theory of "latent structure" (44) will not be discussed here.

In my own work the attention to vicarious functioning has concentrated on perception rather than on behavioral expression (and, within perception, on physical objects or on the more static traits of social objects), and the implications of vicariousness upon representative design have been developed through this medium (7, ch. v, vi, viii; for more general discussion see 8, ch. ii). In his contribution to the present symposium Hammond (20) has expanded this to demand allowance of multiple mediation patterns for, and an attendant multiple cue analysis of, clinical judgment by recourse to the requirements of representative design.

Ecological validity and utilization of clinical cues. A problem ensuing from the work of Frenkel-Brunswik and of Hammond, just cited, is the question about the extent to which the ecological validity of the potential cues available to clinical intuition is duplicated by the weight given these cues in their utilization on the part of the responding clinician. Hammond and his collaborators at Colorado (20), and, independently, Smedslund at Oslo (in an as yet unpublished study), have recently found cases of gross discrepancy in this respect, involving both the ignoring or even the reversal of valid cues and the overutilization of cues of low validity.

Pitfalls of systematic design in biological application. As we have seen, representative design is especially indispensable whenever the relative contribution of different variables in a functional context is the subject of investigation. But it must not be forgotten that any systematic experiment, regardless of how oddly conceived it may be, represents at least one actual or potential ecological instance, and in this sense is a bit of reality; in addition, as we have seen, it may be the only means of obtaining knowledge in the reductive context of science.

Impressive as are the achievements of reductive experimentation in the biological sciences, the dangers resulting from the tying of variables and from other characteristics of nonrepresentative designs have rarely been completely avoided. The history of science furnishes ample evidence of harmful effects of systematic design upon practice or theoretical outlook. As a layman, one may think of hygiene in medicine with its dramatic changes between the complete neglect and the excesses of aseptic or antiseptic policies; the boiling of milk and devitaminization; anemia and the eating of liver, and so forth. All of these practices appear as playballs of variables arbitrarily selected for study, which thus acquire undue prominence and throw the picture of the interplay of factors out of balance; and the shortcuts involved in the ecological overgeneralization of results may be even more serious. The biochemist or nutritionist who shuns eggs but is a chain smoker, and the cancer specialist who does the opposite, are too familiar examples to require elaboration.

Systematic design in cliché literature. We may also regard the "world" of popular novels and movies as an artificial, cultural subecology. Its outstanding feature is the presence of clichés; these include both personality stereotypes and plot formulas. Clichés are similar to experiments using systematic design. Factors that show some degree of independent variability in real life are artificially tied. Old-fashioned opera plots or soap melodramas that "drip with generosity" share with cowboy movies what David Hume would have called "inseparable" associations of noble character, overpowering strength, fairness, courage, youth, final success, and so forth, or of their opposites. The cliché is a "worn" case or incident, by no means impossible or nonexistent but made prominent out of all proportion to its frequency, and to the detriment of all other types of incident. Like the systematic design of experiments, it superimposes artificial "laws" upon an ecology which it thereby depletes.

In the case of the systematic experiment it is these artificial laws in the design which, as we have suggested, are at least in part responsible for the often striking lawfulness of the results. At this point the suspicion arises that the didactic role which systematic experimentation obviously plays in the mental economy of the scientist, by virtue of the simplicity and order it both requires in the design and furnishes in the result, may outweigh the fact-finding competence of systematically designed experiments. Certainly the more drastically simplified forms of art of which we have spoken, and which are so similar in pattern to the systematic experiment, are clearly didactic rather than informative in any realistic sense; we do not go to the movies to find out about life or to form a scientifically airtight theory about personality, in spite of the fact that movies may sometimes be helpful in temporarily smoothing the perplexities of life.

The main function both of art and of systematic experimentation, then, is to shake and mold us by exaggeration and extreme correlation or absence of correlation. But exaggeration is distortion, and this distortion must in science eventually be resolved by allowing the more palatable systematic design to mature into, and to be superseded by, the more truthful representative design.

CONCLUSION: UNITY OF SCIENCE AND UNITY OF PSYCHOLOGY

Our considerations in comparative methodology have brought into focus the thematic diversification that is possible within the over-all unity of exact, observationally physicalistic science. They may facilitate the cheerful relinquishment of the overheated nomothetic bias under which the development of psychology has long suffered and which is making the establishment of a molar-functional psychology an uphill battle. Acceptance of the probabilistic conception, both for the propaedeutic study of ecology and for the functional analysis of perceptual or behavioral achievement, not only sets normative psychology on the right track; it also unites it methodologically with differential psychology, thus lending ideological support to the inherently statistical character they both share, and making a virtue of this often bemoaned necessity.

Psychology thus acquires the distinctive, well-circumscribed internal unity and coherence which we have long searched for, and the reality of which many of us have doubted. This unity emerges, as does perhaps all good unity of—and in—science, as one of methodology or grand design. In both of its major aspects, representative design and functional theory, it centers about probabilistic texture and the ways of its exploitation by the organism. The fact that organismic achievement and its reconstruction in psychology are characterized by certain similarities with predictive procedures in economics, meteorology, and especially in the study of telecommunication through uncontrolled, noisy external media, should help to bring into focus the family of sciences in its entirety. Such a focusing should help to set in proper perspective our traditional thematic dependence on, or need for succorance from, a select few of the older disciplines which may have appeared particularly glamorous, especially physiology and physics.

REFERENCES

1. ATTNEAVE, F. Some informational aspects of visual perception. *Psychol. Rev.*, 1954, **61**, 183–193.
2. BRUNSWIK, E. *Wahrnehmung und Gegenstandswelt.* Vienna: Deuticke, 1934.
3. BBUNSWIK, E. Psychology as a science of objective relations. *Phil. Sci.*, 1937, **4**, 227–260. (Errata corrected 1938, **5**, 110.)
4. BRUNSWIK, E. Probability as a determiner of rat behavior. *J. exp. Psychol.*, 1939, **25**, 175–197.
5. BRUNSWIK, E. Organismic achievement and environmental probability. *Psychol. Rev.*, 1943, **50**, 255–272.
6. BRUNSWIK, E. Distal focussing of perception: size-constancy in a representative sample of situations. *Psychol. Monogr.*, 1944, **56**, No. 1 (Whole No. 254).
7. BRUNSWIK, E. *Systematic and representative design of psychological experiments.* Berkeley: Univer. of California Press, 1947. (Also in J. Neyman [Ed.], *Berkeley symposium on mathematical statistics and probability.* Berkeley: Univer. of California Press, 1949.)
8. BRUNSWIK, E. *The conceptual framework of psychology.* Chicago: Univer. of Chicago Press, 1952. (*Int. Encycl. unified Sci.*, Vol. I, No. 10.)
9. BRUNSWIK, E. "Ratiomorphic" models of perception and thinking. In N. Mailloux (Ed.), *Proc. 14th Int. Congr. Psychol., Montreal, 1954,* in press.
10. BRUNSWIK, E., & HERMA, H. Probability learning of perceptual cues in the establishment of a weight illusion. *J. exp. Psychol.*, 1951, **41**, 281–290.
11. BRUNSWIK, E., & REITER, L. Eindrucks-Charaktere schematisierter Gesichter. *Z. Psychol.*, 1937, **142**, 67–134.
12. DUKES, W. F. Ecological representativeness in studying perceptual size-constancy in childhood. *Amer. J. Psychol.*, 1951, **64**, 87–93.
13. ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
14. FISHER, R. A. *Design of experiments.* Edinburgh: Oliver & Boyd, 1935.
15. FRENKEL-BRUNSWIK, ELSE. Motivation and behavior. *Genet. Psychol. Monogr.*, 1942, **26**, 131–265.
16. GIBSON, J. J. Motion picture testing and research. Washington, D. C.: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No. 7.)
17. GRAHAM, C. H. Visual perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 868–920.
18. HAMMOND, K. R. Subject and object sampling: a note. *Psychol. Bull.*, 1948, **45**, 530–533.

19. HAMMOND, K. R. Representative vs. systematic design in clinical psychology. *Psychol. Bull.*, 1954, **51**, 150–159.

20. HAMMOND, K. R. Probabilistic functioning and the clinical method. *Psychol. Rev.*, in press.

21. HEIDER, F. Environmental determinants in psychological theories. *Psychol. Rev.*, 1939, **46**, 383–410.

22. HEMPEL, C. G. *Fundamentals of concept formation in empirical science.* Chicago: Univer. of Chicago Press, 1952. (*Int. Encycl. unified Sci.*, Vol. II, No. 7.)

23. HOLADAY, B. E. Die Grössenkonstanz der Sehdinge. *Arch. ges. Psychol.*, 1933, **88**, 419–486.

24. HOLT, E. B. *The Freudian wish.* New York: Holt, 1915.

25. HOLWAY, A. H., & BORING, E. G. Determinants of apparent visual size with distance variant. *Amer. J. Psychol.*, 1941, **54**, 21–37.

26. HULL, C. L. *Principles of behavior: an introduction to behavior theory.* New York: D. Appleton-Century, 1943.

27. ITTELSON, W. H. *The Ames demonstrations in perception.* Princeton: Princeton Univer. Press, 1952.

28. JOYNSON, R. B. The problem of size and distance. *Quart. J. exp. Psychol.*, 1949, **1**, 119–135.

29. KATZ, D. *The world of colour.* (Trans. by R. B. McLeod from the 2nd German Ed. [1930].) London: Kegan Paul, 1935. (First German edition, 1911.)

30. KENDALL, M. G. *Advanced theory of statistics.* Vol. II. (3rd Ed.) New York: Hafner, 1951.

31. LANDAHL, H. D., MCCULLOCH, W. S., & PITTS, W. A statistical consequence of the logical calculus of nervous nets. *Bull. math. Biophysics*, 1943, **5**, 135–137.

32. LUNEBURG, R. K. *Mathematical analysis of binocular vision.* Princeton: Princeton Univer. Press, 1947.

33. MCCULLOCH, W. S., & PITTS, W. A logical calculus of the ideas immanent in nervous activity. *Bull. math. Biophysics*, 1943, **5**, 115–133.

34. MCGILL, W. J. Multivariate transmission of information and its relation to analysis of variance. *Hum. Fact. Operat. Res. Lab.*, 1953, Rep. No. 32.

35. MILLER, G. A., & FRICK, F. C. Statistical behavioristics and sequences of responses. *Psychol. Rev.*, 1949, **56**, 311–324.

36. MISES, R. v. *Probability, statistics and truth.* New York: Macmillan, 1939.

37. MOHRMANN, K. Lautheitskonstanz im Entfernungswechsel. *Z. Psychol.*, 1939, **145**, 145–199.

38. RAPOPORT, A. Cycle distributions in random nets. *Bull. math. Biophysics*, 1948, **10**, 145–157.

39. RASHEVSKY, N. *Mathematical biophysics.* (Rev. Ed.) Chicago: Univer. of Chicago Press, 1948.

40. SANDER, F. Experimentelle Ergebnisse der Gestaltpsychologie. *10th Kongr. exp. Psychol., Bonn*, 1928, 23–88.

41. SHANNON, C. E., & WEAVER, W. *Mathematical theory of communication.* Urbana: Univer. of Illinois Press, 1949.

42. SHIMBEL, A., & RAPOPORT, A. Statistical approach to the theory of the nervous system. *Bull. math. Biophysics*, 1948, **10**, 41–55.

43. SKINNER, B. F. The experimental analysis of behavior. In *Proc. 13th Int. Congr. Psychol., 1951.* Stockholm: Lagerström, 1952. Pp. 62–91. (Also in *Amer. Psychologist*, 1953, **8**, 69–78.)

44. STOUFFER, S. A., *et al. Measurement and prediction.* Princeton: Princeton Univer. Press, 1950.

45. THOMSON, A. L. *Bird migration.* (Rev. Ed.) London: Witherby, 1942.

46. THURSTONE, L. L. *A factorial study of perception.* Chicago: Univer. of Chicago Press, 1944.

47. TOLMAN, E. C. *Purposive behavior in animals and men.* New York: Century, 1932.

48. TOLMAN, E. C., & BRUNSWIK, E. The organism and the causal texture of the environment. *Psychol. Rev.*, 1935, **42**, 43–77.

49. VON NEUMANN, J., & MORGENSTERN, O. *Theory of games and economic behavior.* (2nd Ed.) Princeton: Princeton Univer. Press, 1947.

50. WATSON, J. B., & LASHLEY, K. S. *Homing and related activities of birds.* Washington: Carnegie Institution, 1915.

51. WIENER, N. *Extrapolation, interpolation, and smoothing of stationary time series.* New York: Wiley, 1949.

52. WINDELBAND, W. *Geschichte und Naturwissenschaft.* Strassburg: Heitz, 1894.

53. WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938. (2nd Ed., with H. Schlosberg, New York: Holt, 1954.)

Psychological Review
Vol. 62, No. 3, 1955

THE PROBABILITY APPROACH AND NOMOTHETIC THEORY

LEO POSTMAN

University of California

Brunswik has made an eloquent case for the formulation of psychological laws in terms of probability linkages between environmental events and organismic achievements (1, 2, 4). There are three main parts to his thesis. First, certain assumptions are made concerning the present and future aims of psychological description and theory; second, the inadequacies of the nomothetic or "uniformity" point of view for attaining these objectives are pointed out; third, the constructive program of the probability approach and the associated method of representative design are presented. In this paper, we shall consider each of the opposed positions, in turn. The strengths and weaknesses of the probability approach and of the nomothetic point of view will be juxtaposed. While recognizing the cogency of many of Brunswik's criticisms, we shall argue for a reaffirmation of the nomothetic position.

FOCUS ON CONDITIONS REPRESENTATIVE OF REAL LIFE

The aim of psychology, from the probability point of view, is to throw light on behavior as it occurs under natural conditions. "Psychology has to focus its descriptions on what the organisms have become focused on" (4, p. 136). The order of ideas must be the same as the order of things. Many psychologists today would probably be willing to subscribe to this programmatic statement; research should approach real-life conditions as closely as possible. But perhaps they would be tempted to subscribe to this objective more readily than they should. It is not necessarily true that investigations are most profitably carried out under the very conditions which they are intended to explain. We are all familiar with the well-worn examples of engineering problems being solved through theoretical physics and pure mathematics. Such, to be sure, is not always the course of events. Many applied problems can be solved only in applied contexts. Generalizations from restricted laboratory findings to specific life situations must always remain precarious. There are, however, great advantages in entering the concrete situation armed with conclusions obtained under controlled, even though artificial, conditions. As Hilgard has suggested, "*A principle once discovered in a better controlled situation can be validated in a less well controlled one.* Even though the sole aim of learning experimentation were practical, it might prove economical to work in the laboratory in order to find the leads worth testing in practice" (7, pp. 358 f.).

Even if we were to agree, then, that the purpose of psychology is to systematize our knowledge of behavior as it occurs in the organism's habitat, it does not follow that the analysis and explanation of such behavior can be pursued only under "natural" conditions. In behavior analysis, as in problem solving, the *detour principle* may provide the optimal road to solution. By turning our backs on the complexities of behavior in the organism's natural habitat we may arrive more speedily at useful tools of analysis than if we allow ourselves to be guided by the apparent lack of uniformity in daily behavior.

From the probability point of view,

laboratory situations artificially produce correlations among variables which are higher than those found under natural conditions. Granted that the stimulus-response correlations as we observe them daily typically fall short of unity, the explanatory premises from which we would hope to derive these facts of behavior need not themselves have the form of contingency or probability statements. To quote Feigl, explanation may be taken primarily to be a *procedure of inference*, and "the necessity which is bestowed upon the facts by their explanation is the *logical necessity* of the *implication* underlying the inference from assumptions to conclusions" (5, pp. 284 f.). Stimulus-response correlations of less than unity should be derivable from explanatory premises which are stated as universals. And the validity of the explanatory premises may usefully be tested in restricted laboratory situations in which artificially high correlations are produced.

EMPHASIS ON ACHIEVEMENT

Closely related to the focus on real-life situations is the emphasis which the probability point of view puts on organismic achievement. The responses of the organism are described and measured in terms of their "functional validity," i.e., with reference to the organism's ability to attain knowledge about the objective environment, "to reach out cognitively into the farther surroundings" (1, p. 135). Thus, the correlation between physical size and perceived size gauges the functional validity of the size judgment, i.e., measures the degree to which the size perception represents a successful adjustment to the environment. Interest in organismic achievement makes correlational analysis the paramount quantitative method of probability psychology. Correlation coefficients measure the extent to which environmental contingencies are properly evaluated by the organism.

"Achievement" and "functional validity" are terms which denote an *evaluation* of the organism's performance. With such evaluations we cannot, of course, quarrel. We may, however, raise the question whether such normative considerations should mold the description and analysis of psychological facts. Should such evaluations not come more properly *after* psychological description and analysis? And they may, of course, implicitly precede such description, inasmuch as they influence the investigator in the choice of his problems. The analysis of behavior gains in flexibility and generality if it is made without reference to adjustive value.

Let me try to illustrate this point with reference to perception, where emphasis on functional validity may result in limitations on the scope of inquiry. In the past few years, there has been considerable interest in the relation between motivation and perception. Experimental evidence has been offered in support of the hypothesis that motivational factors may significantly influence perceptual discrimination. If true, such facts should hardly be startling. Perceptual discrimination is, after all, behavior, and the study of motivational determinants of behavior is one of the major concerns of contemporary psychology. Nevertheless this work, quite apart from the many experimental uncertainties, has given rise to theoretical controversy quite out of proportion to the magnitude of the empirical results. The core of the difficulty lies, I believe, in explicit or implicit assumptions about the functional validity of perception. The function of perception, most of us firmly believe, is to give correct information about the distal environment. Hence, we expect perceptual theory to explain the ways in which such achievements are realized.

Now data are presented indicating that in some situations there may be systematic biases operating against functional validity. Neither a classical (psychophysical) theory nor a probability theory which is built around the principle of functional validity can readily assimilate such data. On the part of psychophysical theorists there has been a strong tendency to treat such findings, almost by fiat, as nonperceptual, e.g., by assuming that motivation may affect judgment or report but not perception. For example, Gibson (6) attempts to draw a systematic line between "perception" and "misperception," and Pratt (12) argues that only discriminations which show a univocal correlation with dimensions of the physical stimulus should be considered perceptual. Strong opposition has been expressed to the mere possibility that perceptions may be determined by so-called internal conditions of the organism as well as by external stimulus patterns. As might have been expected, the specter of solipsistic encapsulation has been raised (8). The answer to all this is, of course, that we must not prejudge the facts. Motivated perceptual selectivity either does or does not occur. To the extent that it occurs, it is a function of specific and restricted antecedent conditions. How well perceptual discriminations correspond in any given situation to other measurements of the stimulus, i.e., its distal properties, is a further empirical question.

The probability point of view, to be sure, stresses the necessary imperfections of perceptual achievement, due both to the ambiguities in the causal texture of the environment and the limited capacity of the organism to utilize available cues. Variations in motivation would be viewed as contributing to the imperfections of achievement. However, since the problems of perception are formulated primarily in terms of correlations between the distal properties of objects and the responses of the organism, motivational variables are a source of "noise" interfering with the attainment of veridical information. The specific conditions under which such variables are effective and the extent of their contribution do not become the focus of experimental inquiry, because achievement-centered observations are summed over a wide range of motivational conditions.

I have dealt with the motivation-perception problem at some length because it illustrates the restriction upon analysis and the loss of information which we risk when we use achievement as a point of reference for the formulation of our problems.

THE PROBLEMS OF MEDIATION

The uniformity point of view is criticized on the grounds that it concerns itself prematurely and unsuccessfully with mediational processes, i.e., with the specific mechanisms intervening between stimulus and response. There is a tie-up between the nomothetic and microscopic attitudes. As Brunswik says, these associated interests in uniformity and mediation lead to concern with "artificially isolated proximal or peripheral technicalities of mediation" (1, p. 262). From an achievement point of view, principles of mediation established under such conditions are vitiated by the facts of vicarious function: the same achievements can be brought about by a variety of different mechanisms.

The first question I would like to raise is whether there is indeed a necessary connection between the nomothetic position and focus on mediational processes in general, and molecular or microscopic processes in particular. The essence of the nomothetic position is that it looks for uniform laws of behavior, e.g., for stimulus-response correlations which are unity, or, at least, theoretically conceived to be unity. The search for such empirical laws can proceed, at a descriptive level, without concern for mediating processes of any kind, whether molar or molecular. In fact, some of those psychologists who have taken an explicit position against mediational hypotheses are equally firmly in the nomothetic camp. Skinner, for example (even though he defines learning as a change in the probability of response), is a nomothetic theorist who vigorously denies the usefulness of mediational hypotheses. He writes: "A science of behavior must eventually deal with behavior in its relation to certain manipulable variables. Theories—whether neural, mental, or conceptual—talk about intervening steps in these relationships. But instead of prompting us to search for and explore relevant variables, they frequently have quite the opposite effect" (13, p. 194), viz., to create a false sense of security. I do not agree with Skinner, but merely cite him to support the point that the search for nomothetic laws does not necessarily entail a commitment to search for microscopic mediational processes. And certainly the advocates of "dimensional analysis," such as McGeoch (10) and Melton (11), combine a nomothetic with a nontheoretical approach. They have been guided in their experimental research by the dimensions of variation in terms of which learning experiments can be described and measured. Their generalizations take the form of precise, quantitative statements about the dependence of learning on a multitude of experimental conditions. At the same time, they have avoided problems of mediation and have not ventured beyond the assertion of empirical laws.

When nomothetic investigators have gone beyond the statement of empirical laws and have invoked intervening variables, they have not necessarily turned to problems of mediation. Rather, the primary use of intervening variables seems to have been to help in the formulation of empirical laws. In first proposing the use of intervening variables, Tolman made it clear that such constructs were to be used in the specification of empirical relationships which, because of their complexity and because of the multitude of variables involved, defied any simple statement. The function relating behavior to antecedent conditions, he wrote, "is a very complicated function . . . it is in fact so complicated that we at present seem unable to state it in any single simple statement. We find, rather, that we have to handle it by conceiving it as broken down into successive sets of component functions" (16, p. 91). At the same time Tolman was careful to distinguish such operational behaviorism from what he called physiological behaviorism, in which the intervening variables would refer to certain peripheral and central disturbances in the nervous system. Spence takes essentially the same position. Theories employing intervening variables "serve primarily as a device to aid in the formulation of the empirical laws. They consist in guesses as to how the uncontrolled or unknown factors in the system under study are related to the experimentally-known variables" (14, p. 72). In short, many nomothetic theorists, and indeed the most influential ones, have deliberately deferred the problems of microscopic mediation. The behavioral scope of such theories depends on the choice of stimulus and response units to which the intervening variables are anchored. Clearly, the larger the units chosen the less precise will be the predictions which can be made by means of the intervening variables. In that respect, the nomothetic view and the probability view are in the same boat. The difference between them lies in the reasons postulated for the failures of prediction. Proponents of the nomothetic point of view see the difficulty in inadequate knowledge of antecedent conditions, while retaining their belief in the essential uniformity of the behavioral laws (9); proponents of the probability view believe that the laws themselves must be stated in probability terms.

I would conclude that there is no necessary connection between the nomothetic position and focus on microscopic problems of mediation. The nomothetic and probability views do not need to differ with respect to the scope of behavioral problems to which they address themselves. These two points of view do, however, clearly lead to different types of experimental design, and therein I see their most crucial and practically important difference. The problem of representative design which is generated by the probability point of view will be considered next.

REPRESENTATIVE DESIGN

On the methodological side, the probability point of view calls for substitution of representative design for the systematic design which has become traditional in psychology (1). Systematic design is distinguished (*a*) by the orderly manipulation of a limited set of variables, with other variables held constant and (*b*) through practical necessity, by the use of only a restricted range of values of the variables manipulated. Both of these features are criticized as unrepresentative of the conditions which the organism encounters in its natural habitat. Systematic control is said to tie together artificially or to untie variables in ways which violate their natural concomitance. Furthermore, when only a restricted range of values is used, generalizations concerning the effects of a variable become precarious. In the light of these criticisms, representative design calls for the representative sampling of situations or objects as well as subjects. For such a sample, correlations can be determined between distal and/or proximal stimulus characteristics on the one hand, and responses or achievements of the organism on the other. The use of this design necessarily entails the abandonment of classical notions of control; we must take the interaction among variables as we find it in the situations included in the sample. There is, so to speak, controlled noncontrol of the variables.

We must admit that the systematically controlled experiment allows of only limited generalization concerning the effects of the variables manipulated. We know this particularly to be true with regard to motivational variables in human experiments where the experimenter, for practical and ethical reasons, can manipulate only a small fraction of the range of motivational intensities encountered in real-life situations. To a less obvious but probably equally serious degree the same limitations apply to some of the major parameters of learning, such as frequency and the pattern of rewards and punishments. In the face of these difficulties, the value of systematic experimentation can be maintained for several reasons. (*a*) The systematic experiment serves to determine the basic functional relationship between the variables studied, i.e., the shape of the function, which may then be extrapolated to values of the variables not measured in the experiment. If the extrapolation is unsuccessful, further experimentation is indicated. (*b*) The value of systematic experiments may lie not so much in establishing the relationship between the total ranges of two or more variables as in testing the implication of a theoretical proposition. It seems that a large majority of current experiments in the area of learning is conceived for this purpose, often to arbitrate between what are believed to be contradictory implications of the views of Hull and Tolman. The spate of experiments on latent learning is a good example. (*c*) The systematic experimenter may be content for the time being to restrict his generalizations to the conditions studied, to build "miniature systems" preliminary to the attack on the large-scale problems. By and large, the proponents of so-called "dimensional analysis" have adopted this position.

Let us next consider the advantages and disadvantages of representative design in meeting these shortcomings of systematic experiments. That representative design is experimentally fruitful has been abundantly proven by the ingenious studies of Brunswik and his students. Especially noteworthy is the study of "distal focusing in perception" (2), in which a sample of size-judgment situations was obtained in the course of the subject's daily routine. This experiment served to give the principle of size constancy wide situational generality. Perceptual experiments such as these are peculiarly favorable to the application of representative design. With respect to other problems, very serious difficulties may arise.

Clearly, we can proceed with representative sampling only if we have (*a*) knowledge of what the universe to be sampled is, particularly what the units are of which it is composed, and (*b*) if we have identified the dimensions or features with respect to which the sampling is to be representative. In an experiment on size constancy, these requirements can be met without too much difficulty. The universe to be sampled consists of size judgments of specific objects at specific moments in time. The parameters with respect to which we would want the sample to be representative are, among others, the physical and retinal sizes and distances of the objects. It is these variables which have entered into the computation of the correlation coefficients, in the study already cited, to determine the functional validity of the size judgments.

We are confronted with very serious difficulties, however, when we contemplate extension of such a design to other areas of behavior. Take learning as an example. If we were to embark on the enterprise of obtaining a representative sample of learning behaviors, the first question we would face would be what a unit of learning is. How do we identify an instance of learning when we encounter it? I suppose we could arrive at some criteria, but these would inevitably reflect considerations derived from nomothetic theory. We could, for example, look for the appearance of new motor movements and identify each of these as a unit of learning. Or we could define units of learning as new-means-end relationships, in Tolman's sense (15). Our choice of unit will critically affect the nature of the sample we would obtain, to say nothing of the correlations between stimuli, responses, and environmental effects. I am wondering, therefore, whether it is possible to carry out representative designs which are not, in fact, large-scale nomothetic experiments, except for wider sampling of interactions among variables. Such at least appears to be the case when we go beyond the perceptual paradigm, but upon further reflection we may even want to include the perceptual studies under this statement. For the choice of the variables of physical size, retinal size, and judged size implies nomothetic concepts concerning the conditions of size discrimination.

A final word concerning the noncontrol of variables in representative design experiments. The fundamental reason for eliminating systematic controls is, of course, the desire not to interfere with the free play of variables as it occurs naturally. One consequence of this approach is that we may deliberately forego the possibility of measuring precisely the nature and amount of interaction among any given limited set of variables. For example, if color is allowed to vary at random when we sample a variety of physical sizes, it may not be possible to ascertain the exact nature of the interaction between physical size and color in determining perceived size. This is as it should be if we are concerned only with the cue value of physical size over the total range of environmental conditions. We are, however, running a calculated risk of sacrificing potentially valuable information. Suppose it were true that differences in color are important when the physical size is small but make no difference when the physical size is large. Since the color variable would not have been systematically controlled, the combinations of sizes and colors necessary for establishing this interaction may not be present in our sample of objects. Moreover, the over-all correlation between color and size judgments would be low, even though it might be high over a restricted range of physical sizes, and low over the rest. Since the nature and amount of interaction among variables are not measured, it would be difficult to apply the findings of representative design experiments to the prediction or control of specific situations.

CONCLUSION

I would conclude, then, that nomothetic theory can meet the very important and serious criticisms advanced by the probability position along several lines: (*a*) some of the fundamental assumptions of the probability position, particularly the focus on real-life situations and achievement, are open to debate; (*b*) the multiple causation of behavior and uncertainty of mediational processes do not preclude the usefulness of explanatory premises asserting universals; (*c*) both systematic and representative design have a price to pay in terms of sacrifice of information—systematic design by artificially isolating, tying, and untying variables; representative design by foregoing the measurement of the interaction of variables in specific situations. Moreover, representative design in the last analysis must start with nomothetic hypotheses in the selection of variables to be sampled and correlated.

In summary, then, I would like to put in a plea for the *status quo*. However, there is room for peaceful coexistence of the two approaches. The search for nomothetic laws of behavior cannot help but benefit from the probabilist's reminder that behavior does not begin nor end under the controlled conditions of the experimental laboratory.

REFERENCES

1. BRUNSWIK, E. Organismic achievement and environmental probability. *Psychol. Rev.*, 1943, **50**, 255–272.
2. BRUNSWIK, E. Distal focussing of perception. *Psychol. Monogr.*, 1944, **56**, No. 1 (Whole No. 254).
3. BRUNSWIK, E. *Systematic and representative design of psychological experiments.* Berkeley: Univer. of California Press, 1947.
4. BRUNSWIK, E. The conceptual focus of systems. In M. H. Marx (Ed.), *Psychological theory.* New York: Macmillan, 1951. Pp. 131–143.
5. FEIGL, H. Rejoinders and second thoughts (Symposium on operationism). *Psychol. Rev.*, 1945, **52**, 284–288.
6. GIBSON, J. J. Theories of perception. In W. Dennis (Ed.), *Current trends in psychological theory.* Pittsburgh: Univer. of Pittsburgh Press, 1951. Pp. 85–110.

7. HILGARD, E. R. *Theories of learning.* New York: Appleton-Century-Crofts, 1948.

8. HOWIE, D. Perceptual defense. *Psychol. Rev.*, 1952, **59**, 308–315.

9. HULL, C. L. The problem of intervening variables in molar behavior theory. *Psychol. Rev.*, 1943, **50**, 273–291.

10. MCGEOCH, J. A. The vertical dimensions of mind. *Psychol. Rev.*, 1936, **43**, 107–129.

11. MELTON, A. W. Learning. In W. S. Monroe (Ed.), *Encyclopedia of educational research.* New York: Macmillan, 1941. Pp. 667–686.

12. PRATT, C. C. The role of past experience in visual perception. *J. Psychol.*, 1950, **30**, 85–107.

13. SKINNER, B. F. Are theories of learning necessary? *Psychol. Rev.*, 1950, **57**, 193–216.

14. SPENCE, K. W. The methods and postulates of "behaviorism." *Psychol. Rev.*, 1948, **55**, 67–78.

15. TOLMAN, E. C. *Purposive behavior in animals and men.* New York: Century, 1932.

16. TOLMAN, E. C. Operational behaviorism and current trends in psychology. *Proc. 25th Ann. Celebr. Inaug. Grad. Stud. Univer. So. Calif.*, 1936, 89–103.

Psychological Review
Vol. 62, No. 3, 1955

DISCUSSION OF PROBABILISTIC FUNCTIONALISM

ERNEST R. HILGARD
Stanford University

We all owe a debt to Brunswik (2) for insisting that we should sample the environment as well as the individual if we are to develop a satisfactory science of human behavior. Let me first of all add an example to show that we need to test our generalizations in the widest range of ecologies.

The example is from some unpublished work of Sears and Whiting (4) concerned with the effect of child-rearing practices on personality. Sears set out to study the later consequences of the time of weaning, in line with the Freudian conjecture that weaning is frustrating, and so, if attempted too early, results in later disturbances of behavior. Sears found, contrary to the Freudian hypothesis, that the *earlier* the child was weaned, the less he showed certain types of disturbed or neurotic behavior in later childhood. The relationship was systematic enough to show a degree of lawfulness, the least neurotic children being those who had been fed from a cup at birth, and the amount of disturbance increasing with the postponement of weaning. I am making no effort to describe the study in detail, for I can make my point without more precise specification.

Now Whiting, following the same line of conjecture that Sears followed, and working as an anthropologist together with Sears in the same Institute, assembled data on a wide range of nonliterate cultures. He, too, was able to show a relationship between age of weaning and later disturbed behavior of the kind agreed upon between him and Sears and of the same sort that Sears had studied. But the relationship that he found, while also lawful, was the inverse of that which Sears found: he found that the *later* the weaning, the less disturbed the later behavior. Here is a superficial contradiction. It turned out that the range of weaning ages in our culture is much more restricted than that in primitive cultures. The late weaning studied by Sears was at about a year of age, while the early weaning studied by Whiting was at that age, and the late weaning for Whiting came at three or four years of age. Where the two populations, the nonliterate population and the Western culture, overlapped in weaning age, the amount of neurotic behavior in both studies agreed. I hope that I have made my point clear. The contradiction between the Sears and Whiting findings reduces to a difference in their ecological sampling of weaning ages. I give this illustration in support of Brunswik's caution that we may get into trouble unless we sample the whole range.

I would not go the further step, however, and say that Brunswik's wise suggestions force us to the rejection of the nomothetic approach. I believe that we can incorporate his suggestions without revolutionizing our logic of experimentation.

Let me put this another way. When we talk about scientific method we often lose sight of the substantive problems that the practicing scientist faces. The aim of science is, after all, a rather simple one: to yield new information about things and events and relationships, information that is communicable, verifiable, reasonably economical, and relevant to our needs. Hence the scientist, in order to meet these requirements, must have a feel for his subject matter; he must work in an appropriate scale of time and space. If he is measuring tree trunks, he doesn't use micrometers, and if he's studying the gratings of crystals, he doesn't use yardsticks. Now, the use of a micrometer and a yardstick is logically the same, but the scientist has to adapt his tools to his problems. There are no guarantees that the scientist will do good work simply because he uses one scientific logic or another. His critical methods must be adapted to the pitfalls of his particular kind of experimentation.

If nomothetic, test-tube experiments have their problems, as indicated by Professor Brunswik, so have experiments with representative designs. Some of these difficulties were well described by Postman (3), and I will not review them. I would call attention to a problem that arises in connection with the correlational study of size constancy, as used in Brunswik's investigations (1); this is to point out that representative designs are no more foolproof than the other types of systematic design. In his experiment, the subject was asked to estimate the size of various objects in a natural setting—trees, telephone poles, barns, bookcases, inkwells, objects of a wide range of size and at a wide range of distance. The estimates were rather accurate, as represented by correlation coefficients between physical size (as measured at the surface of the object) and the estimates by the subject. Yet, such high correlations could be obtained with this range of data even with many errors in detail; some errors were as great as 3:1, that is, some object might be reported as 3 times its object size, or ⅓ its object size without distorting the correlation in the .90's.

Suppose, now, the experiment were repeated with a blindfolded subject. He would score pretty high, too, by assigning a size appropriate to the name of the object, that is, a house size, an inkwell size, a book size, etc. His estimates too, like those of the subject in Brunswik's experiment, would be much nearer to object size—that is the distal size—than to proximal photographic size. Of course, he would not have a basis for estimating proximal size, for he would not be told the distance of the object, or even if he would be told the distance of the object, he would have no way of using it. In other words, the main conclusions from Brunswik's research in the perceived size constancy would be obtained with a subject who could not perceive because his eyes were bandaged. This follows because in this experiment there is an artificial tying between known size and perceived size. This kind of error could not occur in the laboratory study of size constancy.

I am not seriously suggesting that Professor Brunswik should have used a blindfolded subject as a control in his experiment, yet the suggestion is not actually farfetched. I am making the point only that critical experimentation is always difficult, and no general methodological solution will take care of all contingencies.

I had the feeling, in listening to the papers, that there may be some confusion in overemphasizing the contrast between nomothetic and probabilistic methods of science, in view of the fact that this contrast is found only in selected illustrations. The contrast is very slight, for example, in the biological study of genetics where a definitely nomothetic gene theory works hand in hand with the statistics of Mendelian theory, say in population genetics. Similarly, some of the solutions of problems in thermodynamics are definitely probabilistic without seriously conflicting with nomothetic laws of energy transformation. Therefore, a nomothetic approach does not require a single track or a microscopic explanation; I think this point has already been made by Postman. And the matter of uncertainty: surely uncertainty does not arise just because of the uncertainty of the environment but also in many other ways, so that the distinction between lawfulness and statistics is not unique to the distinction between systematic experimentation and representative design.

Now the matter of working backwards from after-analysis. This bothers me in view of certain very serious problems of partial correlational analysis which I am not going into; but let me simply say that many of the analyses we have had of the nature-nurture controversy by means of correlational analysis lead me to suggest that correlation is an instrument of the devil. One has to be *extremely* careful in making any kind of analysis into scientific laws on the basis of correlational analysis unless one already knows the causal determiners which are, in a sense, nomothetic. That is, if you know the nomothetic analysis, then you can specify sizes, etc., in terms of correlation.

And then I wonder also if the use of aftereffects would have worked, say, in the fields of bacteriology or dietetics which Professor Brunswik has mentioned in passing, as he spoke about enriched bread, or chain smokers. I wonder, for example, if a study of recovery from pneumonia would ever have found that those people who recover more rapidly are those who happen to have taken penicillin; or that the correlation between diabetes and insulin would have ever resulted just from a very careful correlational study of diabetes and its recovery without some kind of experimental intrusion in the way of systematic design. It is very easy to pick examples on one side or the other; perhaps it's unfair to do so, but if you can pick examples on both sides then the question of which method is better seems to me a practical one rather than one of basic theory.

And thus I find myself led to a conclusion rather like Postman's, in defense of the *status quo*. Perhaps I can paraphrase my attitude toward the nomothetic-probabilistic design problem in this way: I should wish to be as nomothetic as I can be, while as probabilistic as I have to be. If we were to accept this solution, I think we would have learned the major lessons to be taught us by Brunswik.

REFERENCES

1. BRUNSWIK, E. Distal focussing of perception: size-constancy in a representative sample of situations. *Psychol. Monogr.*, 1944, 56, No. 1 (Whole No. 254).
2. BRUNSWIK, E. Representative design and probabilistic theory in a functional psychology. *Psychol. Rev.*, 1955, 62, 193–217.
3. POSTMAN, L. The probability approach and nomothetic theory. *Psychol. Rev.*, 1955, 62, 218–225.
4. WHITING, J. W. M. The cross-cultural method. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. I. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 523–531.

Psychological Review
Vol. 62, No. 3, 1955

DISCUSSION: THEORY AND REDUCTIONISM

DAVID KRECH

University of California

I feel that the differences among us are primarily differences in preferences and taste, or perhaps of practice, or of dogma. Brunswik, to use a cliché of Einstein, thinks that "God *does* gamble." I think He does not. Neither of us has any access to God, neither of us knows, and so it's just a matter of preference. The difference between Postman's position and my position is that he does not want to bite off more than he can chew, and I want to bite off everything in one big bite, and that again is a difference of taste.

In the face of Brunswik's persuasive arguments (1), Postman (6) has, in my opinion, countered each one with an equally persuasive counterargument. Point for point (except for one or two rather important ones) I would like to associate myself with those counterarguments. My most general (and perhaps least fair) reaction to Brunswik's position may be summarized by saying that he has taken a methodological *criticism* and made of it a *cosmology*.

I agree with him when he says, "The chief objection against any systematic design and the laws derived from it is that these entail definite dangers of interpretation." Of course they do! And we should be eternally grateful to Brunswik for alerting us to some of the dangers and misinterpretations. But interpretations have been, are, and always will be dangerous—I might say, as satisfying as they are dangerous. Just because there exist dangers of interpretation in the seeking after nomothetic laws is no argument at all for intellectual resignation. We must avoid throwing out the baby with the dirty bath water—even though the experimental water be as murky as Brunswik believes and even though our psychological baby be as mewling and puking as our critics assert. And besides—some of us prefer to interpret dangerously rather than to correlate safely.

But of course Brunswik's championing of probabilistic theory does not derive only from his fears of "definite dangers" of interpretation. He asserts and reasserts that *all relations in psychology remain probabilistic in principle.* And here he loses me completely —or almost completely. Later I shall indicate a possible reinterpretation of Brunswik's dictum with which I have some sympathy.

I will agree that thus far we have no one-to-one relationship in psychology. I will agree that correlations of less than unity can be useful. I will even agree with Brunswik's very acute observation that psychology seems to become more and more probabilistic as it becomes more and more molar. But nevertheless I remain singularly unimpressed with his plea that we must go the whole hog and assert a probabilistic theory. The reasons for my recalcitrance are fundamentally matters of preference and faith. I *prefer* nomothetic laws. I have faith (with Einstein) that unambiguous regularities will be discovered. Let me, therefore, spell out my intuitions (rather than reasons) for disagreeing with Brunswik. And in so doing I think that I can take on Postman at the same time.

Postman, in countering Brunswik's charge that the uniformity point of view concerns itself with *mediational* processes, argued that Brunswik's charge simply was not true. That is, Postman maintained the thesis that there is no necessary tie between the nomothetic and the microscopic attitudes. Here is where I believe Postman nodded, and here is where I believe that Brunswik has the better of the argument. I would not only confess to Brunswik's charge, but I would loudly assert, with complete lack of shame, that this tie is the crowning glory and wonderful promise of the uniformity point of view. In other words, Postman denies a virtue. Postman, in denying this grace, takes a Brunswikian achievement point of view and argues that constructs in psychology are useful only to the extent that they help us in the description and prediction of behavior. And then Postman attacks the neurologizing psychologists because of their unbusinesslike behavior of being too free with promissory notes by quoting Kessen and Kimble's (3) criticism of my own speculations (4) to the effect that "the mere claim that a theory is neurological does precisely nothing to give it greater authority than a purely psychological approach."

And here is what I would have to say to both Brunswik and Postman (remember—I am spelling out my preferences, faiths, and intuitions):

a. I have always made it a cardinal principle to live beyond my income. And although I have yet to find a one-to-one correlation in psychology, Professor Brunswik—or between psychology and neurology, Professor Postman—I am always ready to make another promissory note and promise that if you bear with us we *will* find uniform laws, Professor Brunswik, and we *will* find a neurological basis for behavior, Professor Postman. And if I can't pay off on my first promissory note I will come seeking refinancing. For deficit financing, it seems to me, is a sound economic principle—it certainly is a sound economic principle for any enterprise based on faith; and certainly science and the intellectual enterprise is that. I have faith that despite our repeated and inglorious failures we will someday come to a theory which is able to give a consistent and complete description of reality. But in the meantime, I repeat, you must bear with us.[1]

b. My second point boils down to a simple "Down with Prediction!" Of course we all want to predict. As a scientist I want to encompass the universe. That is, no matter how useful probability statements may be, and no matter how useful one-to-one correlations between stimulus and response may be, I still feel a lack of completion in a correlation coefficient of + 1. Even if a hypothetical construct *does not add one whit* to my predictive power, I would feel much happier for having that construct if it promises to bring into the matrix of correlations *one more observable event,* useless though it may be for the psychologist *qua* psychologist.

c. My third point gives me an opportunity to agree with Brunswik (although on my own terms), and I think that is the proper way to end my comments. I am convinced that it is in the microscopic world that one will most readily find uniform laws. And so Brunswik is eminently right when he argues that psychology becomes more and more probabilistic as it becomes more and more molar. In that sense I understand and agree with Brunswik's dictum that all relations in "psychology" (read "molar psychology") remain probabilistic in nature or in principle. But if that is so, then if we are ever to build *theory* in behavior, we must, as Feigl (2) has expressed it several times, stop being "psychologists" (again read "molar psychologists"). In other words, since molar psychology is doomed to probabilistic relations, then *if theory is our aim* (and it need not be), let us be done with isolationist molar psychology (psychology *qua* psychology, behavior *qua* behavior), let us be done with sociology *qua* sociology, anthropology *qua* anthropology, let us not be frightened by that awful word "reductionism," and let us seek really to become unified scientists by consciously espousing the "microscopic" approach.

---

[1] As I ready these informal comments for publication I cannot refrain from pointing to our recent article on enzyme concentrations in the brain and adjustive behavior patterns (5) as representing, perhaps, a partial token payment on my many outstanding promissory notes.

REFERENCES

1. BRUNSWIK, E. Representative design and probabilistic theory in a functional psychology. *Psychol. Rev.*, 1955, **62**, 193–217.
2. FEIGL, H. Principles and problems of theory construction in psychology. In W. Dennis (Ed.), *Current trends in psychological theory*. Pittsburgh: Pittsburgh Univer. Press, 1951. Pp. 179–213.
3. KESSEN, W., & KIMBLE, G. A. "Dynamic systems" and theory construction. *Psychol. Rev.*, 1952, **59**, 263–267.
4. KRECH, D. Dynamic systems, psychological fields and hypothetical constructs. *Psychol. Rev.*, 1950, **57**, 783–790.
5. KRECH, D., ROSENZWEIG, M. R., BENNETT, E. L., & KRUECKEL, B. Enzyme concentrations in the brain and adjustive behavior-patterns. *Science,* 1954, **120**, 994–996.
6. POSTMAN, L. The probability approach and nomothetic theory. *Psychol. Rev.*, 1955, **62**, 218–225.

Psychological Review
Vol. 62, No. 3, 1955

FUNCTIONALISM, PSYCHOLOGICAL THEORY, AND THE UNITING SCIENCES: SOME DISCUSSION REMARKS

HERBERT FEIGL

University of Minnesota and Minnesota Center for Philosophy of Science

Tolman and Brunswik (1, 24, 25), perhaps more deliberately than any other psychologists in the more recent development of the functionalistic approach in psychology, have emphasized the mutually substitutable nature of the mediating processes which occur between the environmental situation (stimuli, cues, means), and the responses (or achievements) in the behavior of organisms. Brunswik, with sovereign and perspicacious knowledge of the history and the systems of psychology, has illuminated in several publications (see especially 2, 3) the relationships of various ways in which scientific research may "focus" upon the subject matter in the biological-psychological area. The "thematic" and "tactical" differences of these various determinations of the type and level of research are not only compatible with the logical empiricist conception of the unity of science (for an exposition see Carnap, 5; Feigl, 6, 7), but actually support the conception of a unitary science precisely because of the obvious relationships between the assorted possibilities of focusing. It seems to me indeed only a tactical (if not a historical-terminological) question just what type or level of analysis one wishes to reserve for *psychology,* and which other focusings one wishes to delegate to biological ecology, physiology, neurology, or sociology and anthropology.

The mutual substitutibility ("vicarious functioning") of the mediating processes in organisms is indeed a striking feature. Biological thinking has borne its imprint for a long time. There is relatively little in the inorganic parts of nature (but of course a good deal in such inorganic artifacts as the servomechanisms) which invites the sort of concept formation which is so prevalent and fruitful in the bio-psycho-social sciences. Some imaginative scientists (not to mention several excessively imaginative philosophical cosmologists) have perceived principles of organization even in stars and atoms. But while Pauli's exclusion principle might indeed be viewed as basic to chemical organization, and Köhler's physical *Gestalten* (14) may describe a great deal of inorganic and organismic patterning, "vicarious functioning" or "means-ends" relationships seem primarily characteristic of the biological level (and everything above it in the pyramid of the sciences)—and of course in man-made machines.

Since scientific research zealously looks for regularities—be it for the sake of prediction, explanation, or control—it will fasten upon any sort of lawful relationships wherever it can find them. Regularities, however, are of many various forms and types (see Feigl, 10); there are those which we (rather sanguinely) consider *"basic,"* viz., the regularities formulated in the fundamental laws of theoretical physics; and there are other regularities which are obviously dependent on the structure of the systems (e.g., drops, bubbles, crystals, geysers, machines, organisms, social groups, etc.). Phenomena of homeostasis (e.g., regarding the relative stability of the blood temperature or the blood sugar level, etc., in mammals) provide a good example of regularities which depend on the structure (and many functions) of living organisms. The regularities of perception, learning, and motivation as studied in behavioristic psychology likewise depend on the "intact" structure and functioning of the organism. If the "life," i.e., the rise and decline, of fixed stars as studied in astrophysics had as much "teleological" and developmental regularities as does the life of organisms, we would probably have a scientific "ecology" and "theory of adaptation of the stars," as we have ecological, evolutionistic, and developmental biological and psychological disciplines.

Brunswik knows, perhaps better than anyone else, that the subject matter of a science can be "carved out" in a number of ways. Ultimately there is only one criterion by which scientists decide which ways of "focusing" are preferable: "By their fruits ye shall know them." But what sort of fruits are desirable depends on one's interests. Skinner's (22, 23) fruits do not satisfy Krech's (15) appetites—and vice versa. Skinner's interests are focused on prediction-for-the-sake-of-control; Krech's on understanding and explaining. Skinner, like many a positivist before him, has his eye especially on the practical applications of science. The systematic design of his experiments selects certain narrowly circumscribed segments of behavior and examines them under various conditions of drive, stimulation, etc. The laws which he establishes, e.g., regarding response frequencies, are thus highly reliable, though of course statistical, i.e., not strictly deterministic. Brunswik's own research on size constancy in visual perception, to which he referred in the present symposium (4), proceeds by representative sampling (thus achieving a "close-to-the-actual-life-situation character") and thus also results in statistical laws; but it is as little concerned with a detailed theory of the mediating processes as is Skinner's psychology of the "empty organism." The more ambitious attempts of Tolman, Hull, Spence, and others to derive behavior regularities from various systems of theoretical postulates strive for more comprehensive unifications than Skinner or Brunswik in their respective (and different) procedures could ever achieve. But the success of molar behavior theories is still quite problematic, even if for the present it is more distinct than that of neurophysiological theories.

In view of the controversial question of the fruitfulness of the more highly theoretical approaches, one can readily understand why Brunswik tends to limit the subject matter of psychology to the study of certain (statistically certifiable) teleological relations between environmental and achievement variables. There is much in the common life connotations of the term "psychology" which would justify this delimitation of subject matter. But Brunswik, in so restricting his enterprise, cannot plausibly claim to produce a psychological *theory*. As I have argued at some length elsewhere (9), it may well be that sooner or later we shall have to recognize that explanations of the mediating processes can be given fruitfully only in terms of neurophysiology. In order to prevent misunderstandings I wish to say, however, that I should be very happy (and perhaps not too surprised either) if typically *molar* behavior theories were destined to succeed better than they have thus far. After all, classical thermodynamics was, and still is, a very powerful theory—independently of its (even more powerful) molecular interpretation. Here, with some reservations, I share the hopes expressed by Postman (21) and Hilgard (13).

The teleological patterns of perception, learning, and action may lend themselves to explanations of more or less unitary form on the molar level. But an explanation of the very functioning of teleological mechanisms cannot be teleological itself (Nagel, 20); it will no doubt have to be of a relatively micro- or molecular character. This, I take it, is the lesson to be learned from cybernetics. The unification of the sciences which is being achieved along these lines is indeed much more exciting, though admittedly more problematic, than the merely methodological or empirical unity of science. With the abandonment of the earlier (logical positivist) thesis of the *translatability* of scientific theories into observation language, the more ambitious thesis of *unitary* science becomes again more vital and pertinent. The empiricist requirement, as we understand it today, demands only that the nomological net of scientific concepts and propositions be tied to observables in certain points. More accurately, the primitives of scientific theories are given their meaning by implicit definitions, i.e., by the postulates that combine them into a network with one another and with the observables.

Operationism, once helpful in eliminating metaphysical elements from the scientific enterprise, has tended to overstep the limits of its usefulness. There is, and here I agree with Krech (15, 16), in this day and age no longer any reason to be afraid of hypotheses. But (as I wish to remind Krech) it is of course wise to label as "promissory notes," "hopes for theoretical postulates," or "explanation sketches" what in present day neurophysiology can be regarded as no more than a program or blueprint for the derivation of behavior laws. Radical operationism has tended unduly to restrict psychological concept formation to purely "intervening variables" (see MacCorquodale and Meehl, 18), i.e., to dispositional concepts which are merely shorthand for empirical laws. It is questionable whether there are any important concepts at all in psychological theory which can be adequately conceived as pure intervening variables in this sense. (For a suggestive discussion of "conventional constructs," i.e., concepts which are neither pure empirical dispositionals, nor physiological existentials, see Lindzey, 17.) It will not help much either if intervening variables are construed as "open concepts," i.e., as dispositionals introducible by sets of test-condition → test-result conditionals open to additional (partial) specifications of meaning of the same (conditional) form. It is the distinctive mark and advantage of any *theory* with genuine explanatory power that it enables us to *derive* the total set of those conditionals. What would otherwise remain a miscellaneous collection of empirical laws or correlations thus acquires the sort of unity which only a theory can engender. Existential hypotheses or conventional constructs thus play a vital and indispensable role in theory construction; the atomic hypotheses in physics and the neurophysiological hypotheses in the behavior sciences are in many ways analogous to one another as regards their unifying power (see Feigl, 8).

Unitary science, understood in the sense of nomological nets (tied to points of observation), has no sharply separable parts or provinces. Only practical divisions of labor, convenient lines of demarcation of research, can be recognized as separating psychology from biology or physiology on the one side and the social sciences on the other. Even if Brunswik's own predilections in defining the subject matter of his probabilistic-functionalistic psychology prevent him from attaining the theoretical level in his research, he has—in the capacity of a metatheorist (i.e., as a comparative methodologist of science)—provided extremely valuable suggestions regarding the place of psychology in the system of the united (or rather unit*ing*) sciences.

REFERENCES

1. BRUNSWIK, E. *Wahrnehmung und Gegenstandswelt*. Vienna: Deuticke, 1934.
2. BRUNSWIK, E. The conceptual focus of some psychological systems. *J. unified Sci.* (*Erkenntnis*), 1939, **8**, 36–49. (Also in Marx, **19**.)
3. BRUNSWIK, E. *The conceptual framework of psychology*. Chicago: Univer. of Chicago Press, 1952. (*Int. Encycl. unified Sci.*, Vol. I, No. 10.)
4. BRUNSWIK, E. Representative design and probabilistic theory in a functional psychology. *Psychol. Rev.*, 1955, **62**, 193–217.
5. CARNAP, R. *Logical foundations of the unity of science*. Chicago: Univer. of Chicago Press, 1938. (*Int. Encycl. unified Sci.*, Vol. I, No. 1.) (Also in Feigl and Sellars, **12**.)
6. FEIGL, H. Unity of science and unitary science. Paper read at Fifth International Congress for the Unity of Science, Cambridge, Mass., 1939.
7. FEIGL, H. The mind-body problem in the development of logical empiricism. *Rev. int. Phil.*, 1950, **4**. (Also in Feigl and Brodbeck, **11**.)
8. FEIGL, H. Existential hypotheses. *Phil. Sci.*, 1950, **17**, 35–62.
9. FEIGL, H. Principles and problems of theory construction in psychology. In W. Dennis (Ed.), *Current trends in psychological theory*. Pittsburgh: Univer. of Pittsburgh Press, 1951. Pp. 179–213.
10. FEIGL, H. Notes on causality. In H. Feigl & M. Brodbeck (Eds.), *Readings in the philosophy of science*. New York: Appleton-Century-Crofts, 1953. Pp. 408–418.
11. FEIGL, H., & BRODBECK, M. (Eds.) *Readings in the philosophy of science*. New York: Appleton-Century-Crofts, 1953.
12. FEIGL, H., & SELLARS, W. (Eds.) *Readings in philosophical analysis*. New York: Appleton-Century-Crofts, 1949.
13. HILGARD, E. R. Discussion of probabilistic functionalism. *Psychol. Rev.*, 1955, **62**, 226–228.
14. KÖHLER, W. *Dynamics in psychology*. New York: Liveright, 1940.
15. KRECH, D. Discussion: theory and reductionism. *Psychol. Rev.*, 1955, **62**, 229–231.
16. KRECHEVSKY, I. Hypotheses in rats. *Psychol. Rev.*, 1932, **39**, 516–532.
17. LINDZEY, G. Hypothetical constructs. *Psychiatry*, 1953, **16**, 27–33.
18. MACCORQUODALE, K., & MEEHL, P. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, **55**, 95–107. (Also in Feigl and Brodbeck, **11**.)
19. MARX, M. H. (Ed.) *Psychological theory: contemporary readings*. New York: Macmillan, 1951.
20. NAGEL, E. Teleological explanation and teleological systems. In S. Ratner (Ed.), *Vision and action: essays in honor of Horace Kallen*. Rutgers Univer. Press, 1953. (Also in Feigl and Brodbeck, **11**.)
21. POSTMAN, L. The probability approach and nomothetic theory. *Psychol. Rev.*, 1955, **62**, 218–225.
22. SKINNER, B. F. *The behavior of organisms*. New York: D. Appleton-Century, 1938.
23. SKINNER, B. F. *Science and human behavior*. New York: Macmillan, 1953.
24. TOLMAN, E. C. *Purposive behavior in animals and men*. New York: Century, 1932.
25. TOLMAN, E. C., & BRUNSWIK, E. The organism and the causal texture of the environment. *Psychol. Rev.*, 1935, **42**, 43–77.

Psychological Review
Vol. 62, No. 3, 1955

IN DEFENSE OF PROBABILISTIC FUNCTIONALISM: A REPLY

EGON BRUNSWIK

University of California

There are two aspects of probabilistic functionalism: representative design and functional theory (11). Let me consider the arguments concerning theory first.

1. As a starting point, let me take exception to the use of the term "uniformity point of view" in contradistinction to "probability point of view" and also to the identification of the former with the nomothetic approach with which Postman's cogent analysis (27) begins. Uniformity is a matter of reality, or of beliefs about reality, while the search for laws is a matter of what to admit to representation out of that reality. I suspect that one of the main sources of confusing the existential with the representational in the present context is the otherwise splendid book of readings on contemporary psychological theory by Marx (24) in which papers by Hull (20) and by myself (5) were rechristened and juxtaposed under the above labels by the editor. These papers had originally been part of a symposium in which Lewin (23), for once, joined forces with Hull if only in common defense of the nomothetic point of view against my own plea for probabilism.

The term "uniformity point of view," which I myself have never used, erroneously suggests that being a probabilist excludes the belief in uniformity, perhaps in the way in which some of the quantum physicists—or in quite another way the vitalists—do not believe in uniformity. The allegation which Hull made in the same vein at the beginning of his paper, asserting that I do not believe in the "existence" of laws applicable to behavior, is thus based on a misunderstanding. The same holds for the present remarks of Krech (22). Contrary to what he imputes, I *do* believe that "God does *not* gamble"; I fully realize that the impasse of quantum physics is irrelevant to our case, uncertainty being for all practical purposes wiped out at the level of physical macrolaws—which are still microlaws to us psychologists—and that vitalism is mere dogma.

But the crucial point is that while God may not gamble, animals and humans do, and that they cannot help but to gamble in an ecology that is of essence only partly accessible to their foresight. And although an infinite and omniscient intellect could operate by law and ratiocination alone, as a psychologist even such a being would have to follow the methodological postulate of behavior-research isomorphism (10, p. 25) and operate at the probabilistic level of discourse. As in geography or textural ecology, this does not entail the negation of law; it merely entails relinquishing attention to law within certain contexts. As these disciplines do not go "beyond" law, we do not either; and as they are sciences, so is probabilistic psychology.

As we all recognize, I presume, not only the legitimacy but also the necessity of the strictly descriptive or statistical disciplines in addition to the nomothetic sciences, it would seem that Krech's as well as Feigl's (15) tendency to relegate the definition of the basic thema of a science to personal preference and taste amounts to a shrinking away from the tasks and obligations inherent in the scientific potential of reality. Each level of parsimony and of generalization, from the most disjointedly enumerative description to the most universal of laws, entails a challenge of its own. As there must be a balance of occupations in a society so that all the potential demands are filled, so in science at large there must be an equitable distribution among the variety of potential themata. Insistence on reduction as a universal goal of science can only result in blighted spots on the landmap of scientific enterprise. Rather than excommunicating functional psychology from the sciences, advocates of this exclusive form of reductionism—be it persons or groups or even entire phases of the history of our science—are embarking on a self-excommunication from psychology.

Probabilistic functionalism is not, in turn, hostile to reduction. It merely places correlational achievement mapping of generalized functional arcs at the top of a hierarchical pyramid (10, ch. ii); this is followed by the macro-mediational analysis of vicarious attainment strategy. As to the third level in this pyramid, the reductive study of micro-mediational tactics, I agree with Professor Hilgard (19) that this and the entire nomothetic approach should not be "replaced" by the probabilistic approach of the two top levels, although I would like to place reductionism in a marginal position to psychology unless it is executed in firm contact with the two functional aims. In order to reduce, we must know what to reduce. We must reduce "from above," that is, starting from such high-complexity functional units as the lens model. Bertalanffy (1, 2) has in effect done so when he attempted a reductive explanation of "equifinality" in terms of Prigogine's new open-system thermodynamics (see also 10, pp. 42 f.). Most of the "laws" that play a part, say, in learning theory were gathered at a surface level of complexity, and either in oblivion of, or in but casual or mysterious contact with, the more broadly functional contexts in which they are embedded.

The nomothetic-reductionist-systematic type of approach has in the past been overstressed at the expense of the probabilistic-functional-representative approach, notwithstanding all the epoch-making discoveries to which Hilgard has referred. We have had all of the former and nothing of the latter for too long. Now we must balance psychology in the molar and molecular realm.

In contrast with most previous movements in psychology, probabilistic functionalism thus does not exclude anything; rather, its scope is encompassing enough to allow us to envisage a system to end all systems.

2. We have recently been admonished by Koch (21) to expand the use of the label "theory" beyond the nomothetic endeavors in the narrower sense. What Koch, and probably Feigl in the present symposium, mean by "metatheory" is theory about theorizing, perhaps including matters of differential thema and design as subsumable under the label of comparative methodology.

But it seems to me that the liberalization of our notions of theory must go still further. We must include within the scope of theory proper the substantive part of functional considerations also, involving primarily the structure of the lens model as a ratiomorphic explication of achievement rather than merely its reduction to physiology or physics. This model involves focal points, areas of unspecificity, mechanisms of substitution, and other devices of multiple mediation. When Feigl demands generality of a theory, the lens model has it: in fact, the very essence of generalized achievement is incorporated within it, and it can be transferred from perception to action and other psychological functions. It even may be used for "explanation" in the sense of subsuming a single act under a generalized "wish," ascertained by the probabilistic application of operational criteria of "purposiveness" such as those of Tolman (see 29, ch. i). The lens model also has productivity: predictions concerning the efficiency and foolproofness of functional attainment, concerning performance in other areas or dimensions of the ecology, and so forth, may be "derived" from it, and the probabilistic character of these derivations should not detract from the potential rigor and objectivity of the procedures involved.

In other words, we must restore the term "model" to its original meaning, as in speaking of a model of San Francisco Bay water currents, or of the models of medieval technologists. In speaking of "models," many psychologists have come to confine themselves to what essentially amounts to mathematical curve fitting in the tradition of strictly nomothetic theorizing; as we have seen, however, this not only is a labor of Sisyphus but also a labor the fruits of which may be blown over like a house of cards when representative design is applied.

Professor Feigl's reluctance to part with the cherished technical meaning of the term "theory" is, to be sure, reinforced by the majority trend in recent psychological "theory construction." But we must insist that the opening up of probabilistic vistas requires some flexibility, some of the "positivism, not negativism" which he himself has preached in other contexts (13) but which, as it must seem from his remarks, he has not always exercized when the challenge was to recognize the growth potential of the psychological thema. This technical meaning of theory, incorporated as it is in Feigl's hierarchy of "levels of explanation," which stretches from strict description through surface laws to low-order and high-order theory (14), reveals only too clearly its origin in the situation in physics with its reduction of, say, the low-order theory of planetary motion by Kepler to the higher-order theory of gravitation by Newton and eventually to the theory of Einstein. Trying to enforce this content-bound schema on another domain is but one more instance of thematic physicalism. Like all automatic transfer of content-alien scaffoldings, it can only result in obscuring the indigenous thema and theoretical structure of the host discipline.

That the question of the applicability of the term "theory" cannot be dismissed as a mere matter of terminology became clear to me when I was asked, at Gustav Bergmann's instigation, to write a note on one of the less well-known papers by Hammond, one in which the situation created in psychology by representative design was set in analogy to that in relativity physics, with special reference to the fact that in both cases regularities hitherto considered universal turn out to be contingent upon a limited ecology (18). Bergmann saw a difficulty in the fact that it was a "design" that was juxtaposed to a "theory." In my reply to this argument (9) I pointed out that the methodology of representative design was inseparable from an inherent theory, and that this theory had been made sufficiently explicit in the original publication (7) under the more modest label of "generalization." The same policy has in part been followed in my main contribution to the present symposium (11).

3. If I understand correctly, Krech agrees with the general thesis that the nomothetic and the microscopic approach go together, while Postman does not. The latter correctly points to Skinner as a psychologist who is nomothetic yet nonmediational. But micromediationism, reductionism, or physiologism are but a few of many possible forms of being atomistic. Another is the encapsulation within the boundaries of the organism, of which Skinner is equally guilty with Hull, and which is

the price paid for some success along nomothetic lines.

It also must be held against Postman that only if our aim is microexplanation or nomothesis at large must we depart in our investigation from the conditions which we intend to explain. The same holds with respect to what Feigl has pointed out in referring to Nagel's telling objection that explanations of teleology cannot in themselves be teleological. The aim of functional psychology is indeed like that of applied psychology, as Postman correctly imputes; or perhaps it goes even further than applied psychology, which has employed systematic design against its own best interests.

4. And now to Hilgard's penetrating argument on representative design as it affects our sampling study of size constancy (6). His suggestion of reading the names of the objects to a blindfolded subject and obtaining estimates of size on this basis will be followed through. Let me venture a guess as to the outcome. In our ecology the functional achievement correlations will be still high, although appreciably lower than in the original study with its unhampered use of the common direct cues. But it is not true that with the use of words we are necessarily overstepping the boundaries of perception in an uncontrollable way, as Hilgard seems to imply. On the contrary, perception must be seen as a cross-departmentally unitary function, and words or other auditory cues are as legitimate a part of the potential mediating cue family hierarchy as are the visual—and in fact the tactile-kinesthetic—cues. Since object names had to be used by the subject of the original study in her communication with the recorder in the course of the observations, Hilgard's suggested design falls more in the category of impoverished rather than in that of shifted cue situations. Under fully representative design, it would be a violation of basic policy *not* to include name cues and the analysis of their effects, rather than the contrary as Hilgard seems to assume.[1]

One must not forget that, ideally, representative design would require huge, concerted group projects, involving as it does a practically limitless number of variables, but that at the same time its rewards in after-the-fact evaluation could be equally widely ramified. In the pilot study under discussion (6) only a select few of these evaluation possibilities were exploited, and the entire problem syndrome of cue participation was left in the dark save for the indirect ascertainment, via the sampling of the objects, of a reasonable degree of representativeness. Since, in principle, under representative design all variables are allowed to vary and none are held constant artificially, their role can be ascertained after the fact. For the same reason, there is a gain rather than a loss of information, contrary to what Postman seems to fear. And we may add that partial correlation, on which Hilgard has expressed reservations, is just one example out of many possible mathematical ways open to, or to be developed by, the statistician for this after-the-fact analysis.

---

[1] As this goes to press I have before me an as yet unpublished, independently conceived study by R. C. Bolles and D. E. Bailey in which a representative sample of $N = 54$ familiar objects was presented to $n = 5$ adult subjects first nonvisually by naming or brief verbal description, and then visually. The authors have kindly given me permission to quote from their results. Nonvisual estimates were almost as accurate ($r = .988$) as visual estimates ($r = .994$), and in fact considerably more accurate than this writer would have guessed offhand. Yet the difference between the two achievements was found to be statistically significant (using the ecological N of 54, as is called for under representative design); in addition, the errors under the two conditions correlated to a relatively moderate extent only. The cue systems involved in the two performances thus seem to a considerable extent nonoverlapping, perhaps having less in common than I conjectured when I suggested above that the verbal situations were merely impoverishments of the full perceptual situations.

Our earlier prediction of the outcome of Hilgard's suggested naming experiment had to be specified to our ecology. Another ecology is conceivable in which the common objects sampled, such as books, windows, trees, hills, and so forth would be so univocally standardized that, given proper perceptual learning, the name could be utilized as a foolproof cue to size. Hilgard's variant of our design would in this case yield perfect correlation, while visual cues could merely upset this achievement. With such a change in ecology the trend of results would thus be reversed, much as in Hilgard's example from Sears and Whiting there was a trend reversal which likewise could be subsumed under representative design.

I should like to add parenthetically that this latter example seems to me to involve primarily a responder-populational variable—weaning age and its range—and only secondarily an ecological impasse; but since both such types of universes should be treated alike and sampled under representative design, this difference is fundamentally beside the point here.

5. A similar expansiveness applies to the case of Postman's concern lest motivational bias be excluded from representative design. Representative design merely complements and completes the existing ways in which the necessity of representative sampling has been recognized, so that in the end such sampling will extend over all psychologically relevant variables.

This does not imply, however, that the concept of functional research should be equally broadly conceived. Postman is correct in assuming that the "veridicality" of perception is the major issue in the study of its functional validity, and that our brand of functionalism is not directly and not primarily concerned with need variables or the structure of the need system. Only insofar as needs or values help the cause of, or interfere with, veridicality—and this may be the case either in "perception" or in "judgment," as Postman correctly remarks—do they become part of the functionalist's direct concern. Even interference may be functionally constructive, such as when needs serve as emphasizers of the size or of some other attribute of a life-relevant object. Furthermore, as we have said (11), values form part of the system of natural-cultural "laws" and as such in themselves become objectives of cool, veridical attainment; it is primarily for this reason, and not because of their entanglement with the world of needs, that we incorporated them in our earlier perceptual constancy program (4). And we may add that all these variables are as much a potential part of an isolating statistical analysis as are Hilgard's word cues.

It is only when it comes to delving into the need structures themselves, their variability and their etiology, as is the case, say, in Bruner and Goodman's comparison of the effect of values on rich vs. poor subjects (3), or in Postman, Bruner, and McGinnies' development of such dynamic aspects as "perceptual defense" (28), that the wisdom of applying the label of functionalism may be doubted.[2] This must be made explicit, especially in view of a somewhat unfortunate current trend toward an overexpansion of the concept of functionalism, by Osgood (23, pp. 294 f.) and by others, according to which all such need considerations would be seen as an outgrowth of the functionalist tradition. Bruner and Postman can themselves not be entirely absolved of having contributed to this trend (for further discussion see 8). In genuinely functional analysis needs appear as mere background elements, only as necessary for all utilitarian considerations—that is, essentially as constant parameters rather than as actual variables or as systems of interdependent variables—the latter being exemplified by Frenkel-Brunswik's "personality-centered" approach to perception (17).

---

[2] *Note by Leo Postman:* As I have tried to show elsewhere (26), the phenomena originally interpreted as perceptual defense can be adequately accounted for in terms of (*a*) principles of associative learning and (*b*) variables governing *S*s' performance in the experimental situation. On these grounds the relevance of such phenomena to functionalist analysis in the narrower sense specified by Brunswik may be defended.

Let me conclude with a consideration of Postman's plea for the *status quo*. As has been pointed out repeatedly in this symposium, much depends on definition. Therefore, if it is being recognized that representative design and the resulting functional theory are a going and a growing concern, while systematic design and the nomothetic approach are about to level off and to reach the point of diminishing returns —if, in other words, the *status quo* is conceived not in a static but in a dynamic manner—then I too am heartily in favor of this *status quo*. And when I hear the acknowledgment, as I think I have by Postman and by Hilgard in the present discourse, that systematic design is a preliminary and an eye-opener to the attack on a larger, representative scale, then I am confident that this *status quo* will indeed be a dynamic one.

The peaceful coexistence of which Postman speaks may even be turned into a situation in which the probabilistic approach begins to repay the debt it owes to the reductionist vanguard. In discussing Hilgard's argument on representative design we have come across the problem of perceptual learning. The presence of such learning must be assumed on the broadest possible basis if perception is to be adjustment to an ecology. It is for this reason that an experimental study of the acquired utilization of artificial illumination cues in perceptual color constancy was made part of our Vienna program; the finding was that such acquisition is possible to a certain extent but that perceptual learning requires slow "stamping in" and is surprisingly autonomous of explicit intellectual awareness (Fieandt, 16).

Even purely ecological surveys may become fruitful for the theory of perception and learning. This writer and Kamiya (12) have demonstrated (with the use of $N = 892$ separations between adjacent parallel lines in a roughly "proportionate" sample of shots from a current motion picture) that the long-recognized gestalt factor of "proximity" possesses a certain modest ($r = .12$) but statistically significant ecological validity as an indicator of mechanical object unity. Its utilization as an organizing principle for perceptual "figures" is thus a probabilistically adjustive mechanism. The realization of this fact may help to open the door for a possible "reduction" of the hitherto unabsorbed gestalt dynamics into learning theory. The latter may thus eventually become universalized under the associationist-functionalist aspect, much as Newton's theory has incorporated both terrestrial gravitation laws and Kepler's laws of planetary motion within a common higher-order theory.

REFERENCES

1. BERTALANFFY, L. v. The theory of open systems in physics and biology. *Science*, 1950, 3, 23–29.
2. BERTALANFFY, L. v. Theoretical models in biology and psychology. *J. Pers.*, 1951, **20**, 24–38.
3. BRUNER, J. S., & GOODMAN, C. C. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33–44.
4. BRUNSWIK, E. *Wahrnehmung und Gegenstandswelt*. Vienna: Deuticke, 1934.
5. BRUNSWIK, E. Organismic achievement and environmental probability. *Psychol. Rev.*, 1943, **50**, 255–272. (Also in Marx, **24**.)
6. BRUNSWIK, E. Distal focussing of perception: size-constancy in a representative sample of situations. *Psychol.*

Monogr., 1944, **56**, No. 1 (Whole No. 254).

7. BRUNSWIK, E. *Systematic and representative design of psychological experiments.* Berkeley: Univer. of California Press, 1947. (Also in J. Neyman [Ed.], *Berkeley symposium on mathematical statistics and probability.* Berkeley: Univer. of California Press, 1949.)
8. BRUNSWIK, E. Remarks on functionalism in perception. *J. Pers.*, 1949, **18**, 56–65.
9. BRUNSWIK, E. Note on Hammond's analogy between "relativity and representativeness." *Phil. Sci.*, 1951, **18**, 212–217.
10. BRUNSWIK, E. *The conceptual framework of psychology.* Chicago: Univer. of Chicago Press, 1952. (*Int. Encycl. Unified Sci.*, Vol. I, No. 10.)
11. BRUNSWIK, E. Representative design and probabilistic theory in a functional psychology. *Psychol. Rev.*, 1955, **62**, 193–217.
12. BRUNSWIK, E., & KAMIYA, J. Ecological cue-validity of 'proximity' and of other Gestalt factors. *Amer. J. Psychol.*, 1953, **66**, 20–32.
13. FEIGL, H. Logical empiricism. In D. Runes (Ed.), *Twentieth century philosophy.* New York: Philosophical Library, 1943. (Also in H. Feigl & W. Sellars [Eds.], *Readings in philosophical analysis.* New York: Appleton-Century-Crofts, 1949. Pp. 3–26.)
14. FEIGL, H. Rejoinders and second thoughts (Symposium on operationism). *Psychol. Rev.*, 1945, 52, 284–288. (Also as: Some remarks on the meaning of scientific explanation. In H. Feigl & W. Sellars [Eds.], *Readings in philosophical analysis.* New York: Appleton-Century-Crofts, 1949. Pp. 510–514.)
15. FEIGL, H. Functionalism, psychological theory, and the uniting sciences: some discussion remarks. *Psychol. Rev.*, 1955, 62, 232–235.
16. FIEANDT, K. v. Dressurversuche an der Farbenwahrnehmung. *Arch. ges. Psychol.*, 1936, **96**, 467–495.
17. FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18**, 108–143.
18. HAMMOND, K. R. Relativity and representativeness. *Phil. Sci.*, 1951, **18**, 208–211.
19. HILGARD, E. R. Discussion of probabilistic functionalism. *Psychol. Rev.*, 1955, **62**, 226–228.
20. HULL, C. L. The problem of intervening variables in molar behavior theory. *Psychol. Rev.*, 1943, **50**, 273–291. (Also in Marx, **24**.)
21. KOCH, S. Theoretical psychology, 1950: an overview. *Psychol. Rev.*, 1951, **58**, 295–301.
22. KRECH, D. Discussion: theory and reductionism. *Psychol. Rev.*, 1955, **62**, 229–231.
23. LEWIN, K. Defining the "field at a given time." *Psychol. Rev.*, 1943, **50**, 292–310. (Also in Marx, **24**.)
24. MARX, M. H. *Psychological theory: contemporary readings.* New York: Macmillan, 1951.
25. OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.
26. POSTMAN, L. On the problem of perceptual defense. *Psychol. Rev.*, 1953, **60**, 298–306.
27. POSTMAN, L. The probability approach and nomothetic theory. *Psychol. Rev.*, 1955, **62**, 218–225.
28. POSTMAN, L., BRUNER, J. S., & MCGINNIES, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, **43**, 142–154.
29. TOLMAN, E. C. *Purposive behavior in animals and men.* New York: Century, 1932.

(Manuscript for this symposium received August 17, 1954)

VOL. 62, NO. 4 JULY, 1955

THE PSYCHOLOGICAL REVIEW

DRIVES AND THE C.N.S. (CONCEPTUAL NERVOUS SYSTEM)[1]

D. O. HEBB

McGill University

The problem of motivation of course lies close to the heart of the general problem of understanding behavior, yet it sometimes seems the least realistically treated topic in the literature. In great part, the difficulty concerns that c.n.s., or "conceptual nervous system," which Skinner disavowed and from whose influence he and others have tried to escape. But the conceptual nervous system of 1930 was evidently like the gin that was being drunk about the same time; it was homemade and none too good, as Skinner pointed out, but it was also habit-forming; and the effort to escape has not really been successful. Prohibition is long past. If we *must* drink we can now get better liquor; likewise, the conceptual nervous system of 1930 is out of date and—if we must neurologize—let us use the best brand of neurology we can find.

Though I personally favor both alcohol and neurologizing, in moderation, the point here does not assume that either is a good thing. The point is that psychology is intoxicating itself with a worse brand than it need use. Many psychologists do not think in terms of neural anatomy; but merely adhering to certain classical frameworks shows the limiting effect of earlier neurologizing. Bergmann (2) has recently said again that it is logically possible to escape the influence. This does not change the fact that, in practice, it has not been done.

Further, as I read Bergmann, I am not sure that he really thinks, deep down, that we should swear off neurologizing entirely, or at least that we should all do so. He has made a strong case for the functional similarity of intervening variable and hypothetical construct, implying that we are dealing more with differences of degree than of kind. The conclusion *I* draw is that both can properly appear in the same theory, using intervening variables to whatever extent is most profitable (as physics for example does), and conversely not being afraid to use some theoretical conception merely because it might become anatomically identifiable.

For many conceptions, at least, MacCorquodale and Meehl's (26) distinction is relative, not absolute; and it must also be observed that physiological psychology makes free use of "dispositional concepts" as well as "existential" ones. Logically, this leaves room for some of us to make more use of explicitly physiological constructs than others, and still lets us stay in communication with one another. It also shows how one's views concerning motivation, for example, might be more influenced than one thinks by earlier physiological notions, since it means that an explicitly physiological conception might be restated in words that have—apparently—no physiological reference.

---

[1] Presidential address, Division 3, at American Psychological Association, New York, September, 1954. The paper incorporates ideas worked out in discussion with fellow students at McGill, especially Dalbir Bindra and Peter Milner, as well as with Leo Postman at California, and it is a pleasure to record my great indebtedness to them.

What I propose, therefore, is to look at motivation as it relates to the c.n.s.—or conceptual nervous system—of three different periods: as it was before 1930, as it was say 10 years ago, and as it is today. I hope to persuade you that some of our current troubles with motivation are due to the c.n.s. of an earlier day, and ask that you look with an open mind at the implications of the current one. Today's physiology suggests new psychological ideas, and I would like to persuade you that they make psychological sense, no matter how they originated. They might even provide common ground—not necessarily agreement, but communication, something nearer to agreement—for people whose views at present may seem completely opposed. While writing this paper I found myself having to make a change in my own theoretical position, as you will see, and though you may not adopt the same position you may be willing to take another look at the evidence, and consider its theoretical import anew.

Before going on it is just as well to be explicit about the use of the terms motivation and drive. "Motivation" refers here in a rather general sense to the energizing of behavior, and especially to the sources of energy in a particular set of responses that keep them temporarily dominant over others and account for continuity and direction in behavior. "Drive" is regarded as a more specific conception about the way in which this occurs: a hypothesis of motivation, which makes the energy a function of a special process distinct from those S-R or cognitive functions that are energized. In some contexts, therefore, "motivation" and "drive" are interchangeable.

MOTIVATION IN THE CLASSICAL (PRE-1930) C.N.S.

The main line of descent of psychological theory, as I have recently tried to show (20), is through associationism and the stimulus-response formulations. Characteristically, stimulus-response theory has treated the animal as more or less inactive unless subjected to special conditions of arousal. These conditions are first, hunger, pain, and sexual excitement; and secondly, stimulation that has become associated with one of these more primitive motivations.

Such views did not originate entirely in the early ideas of nervous function, but certainly were strengthened by them. Early studies of the nerve fiber seemed to show that the cell is inert until something happens to it from outside; therefore, the same would be true of the collection of cells making up the nervous system. From this came the explicit theory of drives. The organism is thought of as like a machine, such as the automobile, in which the steering mechanism—that is, stimulus-response connections—is separate from the power source, or drive. There is, however, this difference: the organism may be endowed with three or more different power plants. Once you start listing separate ones, it is hard to avoid five: hunger, thirst, pain, maternal, and sex drives. By some theorists, these may each be given a low-level steering function also, and indirectly the steering function of drives is much increased by the law of effect. According to the law, habits—steering functions—are acquired only in conjunction with the operation of drives.

Now it is evident that an animal is often active and often learns when there is little or no drive activity of the kinds listed. This fact has been dealt with in two ways. One is to postulate additional drives—activity, exploratory, manipulatory, and so forth. The other is to postulate acquired or learned drives, which obtain their energy, so to speak, from association with primary drives.

It is important to see the difficulties to be met by this kind of formulation, though it should be said at once that I do not have any decisive refutation of it, and other approaches have their difficulties, too.

First, we may overlook the rather large number of forms of behavior in which motivation cannot be reduced to biological drive plus learning. Such behavior is most evident in higher species, and may be forgotten by those who work only with the rat or with restricted segments of the behavior of dog or cat. (I do not suggest that we put human motivation on a different plane from that of animals [7]; what I am saying is that certain peculiarities of motivation increase with phylogenesis, and though most evident in man can be clearly seen with other higher animals.) What is the drive that produces panic in the chimpanzee at the sight of a model of a human head; or fear in some animals, and vicious aggression in others, at the sight of the anesthetized body of a fellow chimpanzee? What about fear of snakes, or the young chimpanzee's terror at the sight of strangers? One can accept the idea that this is "anxiety," but the anxiety, if so, is not based on a prior association of the stimulus object with pain. With the young chimpanzee reared in the nursery of the Yerkes Laboratories, after separation from the mother at birth, one can be certain that the infant has never seen a snake before, and certainly no one has told him about snakes; and one can be sure that a particular infant has never had the opportunity to associate a strange face with pain. Stimulus generalization does not explain fear of strangers, for other stimuli in the same class, namely, the regular attendants, are eagerly welcomed by the infant.

Again, what drive shall we postulate to account for the manifold forms of anger in the chimpanzee that do not derive from frustration objectively defined (22)? How account for the petting behavior of young adolescent chimpanzees, which Nissen (36) has shown is independent of primary sex activity? How deal with the behavior of the female who, bearing her first infant, is terrified at the sight of the baby as it drops from the birth canal, runs away, never sees it again after it has been taken to the nursery for rearing; and who yet, on the birth of a *second* infant, promptly picks it up and violently resists any effort to take it from her?

There is a great deal of behavior, in the higher animal especially, that is at the very best difficult to reduce to hunger, pain, sex, and maternal drives, plus learning. Even for the lower animal it has been clear for some time that we must add an exploratory drive (if we are to think in these terms at all), and presumably the motivational phenomena recently studied by Harlow and his colleagues (16, 17, 10) could also be comprised under such a drive by giving it a little broader specification. The curiosity drive of Berlyne (4) and Thompson and Solomon (46), for example, might be considered to cover both investigatory and manipulatory activities on the one hand, and exploratory, on the other. It would also comprehend the "problem-seeking" behavior recently studied by Mahut and Havelka at McGill (unpublished studies). They have shown that the rat which is offered a short, direct path to food, and a longer, variable and indirect pathway involving a search for food, will very frequently prefer the more difficult, but more "interesting" route.

But even with the addition of a curiosity-investigatory-manipulatory drive, and even apart from the primates, there is still behavior that presents difficulties. There are the reinforcing effects of incomplete copulation (43) and of saccharin intake (42, 11), which do not reduce to secondary reward. We must not multiply drives beyond reason, and at this point one asks whether there is no alternative to the theory in this form. We come, then, to the conceptual nervous system of 1930 to 1950.

MOTIVATION IN THE C.N.S. OF 1930–1950

About 1930 it began to be evident that the nerve cell is not physiologically inert, does not have to be excited from outside in order to discharge (19, p. 8). The nervous system is alive, and living things by their nature are active. With the demonstration of spontaneous activity in c.n.s. it seemed to me that the conception of a drive system or systems was supererogation.

For reasons I shall come to later, this now appears to me to have been an oversimplification; but in 1945 the only problem of motivation, I thought, was to account for the *direction* taken by behavior. From this point of view, hunger or pain might be peculiarly effective in guiding or channeling activity but not needed for its arousal. It was not surprising, from this point of view, to see human beings liking intellectual work, nor to find evidence that an animal might learn something without pressure of pain or hunger.

The energy of response is not in the stimulus. It comes from the food, water, and oxygen ingested by the animal; and the violence of an epileptic convulsion, when brain cells for whatever reason decide to fire in synchrony, bears witness to what the nervous system can do when it likes. This is like a whole powder magazine exploding at once. Ordinary behavior can be thought of as produced by an organized series of much smaller explosions, and so a "self-motivating" c.n.s. might still be a very powerfully motivated one. To me, then, it was astonishing that a critic could refer to mine as a "motivationless" psychology. What I had said in short was that any organized process in the brain is a motivated process, inevitably, inescapably; that the human brain is built to be active, and that as long as it is supplied with adequate nutrition will continue to be active. Brain activity is what determines behavior, and so the only behavioral problem becomes that of accounting for *in*activity.

It was in this conceptual frame that the behavioral picture seemed to negate the notion of drive, as a separate energizer of behavior. A pedagogical experiment reported earlier (18) had been very impressive in its indication that the human liking for work is not a rare phenomenon, but general. All of the 600-odd pupils in a city school, ranging from 6 to 15 years of age, were suddenly informed that they need do no work whatever unless they wanted to, that the punishment for being noisy and interrupting others' work was to be sent to the playground to play, and that the reward for being good was to be allowed to do more work. In these circumstances, *all* of the pupils discovered within a day or two that, within limits, they preferred work to no work (and incidentally learned more arithmetic and so forth than in previous years).

The phenomenon of work for its own sake is familiar enough to all of us, when the timing is controlled by the worker himself, when "work" is not defined as referring alone to activity imposed from without. Intellectual work may take the form of trying to understand what Robert Browning was trying to say (if anything), to discover what it is in Dali's paintings that can interest others, or to predict the outcome of a paperback mystery. We systematically underestimate the human need of intellectual activity, in one form or another, when we overlook the intellectual component in art and in games. Similarly with riddles, puzzles, and the puzzle-like games of strategy such as bridge, chess, and *go;* the frequency with which man has devised such problems for his own solution is a most significant fact concerning human motivation.

It is, however, not necessarily a fact that supports my earlier view, outlined above. It is hard to get these broader aspects of human behavior under laboratory study, and when we do we may expect to have our ideas about them significantly modified. For my views on the problem, this is what has happened with the experiment of Bexton, Heron, and Scott (5). Their work is a long step toward dealing with the realities of motivation in the well-fed, physically comfortable, adult human being, and its results raise a serious difficulty for my own theory. Their subjects were paid handsomely to do nothing, see nothing, hear or touch very little, for 24 hours a day. Primary needs were met, on the whole, very well. The subjects suffered no pain, and were fed on request. It is true that they could not copulate, but at the risk of impugning the virility of Canadian college students I point out that most of them would not have been copulating anyway and were quite used to such long stretches of three or four days without primary sexual satisfaction. The secondary reward, on the other hand, was high: $20 a day plus room and board is more than $7000 a year, far more than a student could earn by other means. The subjects then should be highly motivated to continue the experiment, cheerful and happy to be allowed to contribute to scientific knowledge so painlessly and profitably.

In fact, the subject was well motivated for perhaps four to eight hours and then became increasingly unhappy. He developed a need for stimulation of almost any kind. In the first preliminary exploration, for example, he was allowed to listen to recorded material on request. Some subjects were given a talk for 6-year-old children on the dangers of alcohol. This might be requested, by a grown-up male college student, 15 to 20 times in a 30-hour period. Others were offered, and asked for repeatedly, a recording of an old stock-market report. The subjects looked forward to being tested, but paradoxically tended to find the tests fatiguing when they did arrive. It is hardly necessary to say that the whole situation was rather hard to take, and one subject, in spite of not being in a special state of primary drive arousal in the experiment but in real need of money outside it, gave up the secondary reward of $20 a day to take up a job at hard labor paying $7 or $8 a day.

This experiment is not cited primarily as a difficulty for drive theory, although three months ago that is how I saw it. It *will* make difficulty for such theory if exploratory drive is not recognized; but we have already seen the necessity, on other grounds, of including a sort of exploratory-curiosity-manipulatory drive, which essentially comes down to a tendency to seek varied stimulation. This would on the whole handle very well the motivational phenomena observed by Heron's group.

Instead, I cite their experiment as making essential trouble for my own treatment of motivation (19) as based on the conceptual nervous system of 1930 to 1945. If the thought process is internally organized and motivated, why should it break down in conditions of perceptual isolation, unless emotional disturbance intervenes? But it did break down when no serious emotional change was observed, with problem-solving and intelligence-test performance significantly impaired. Why should the subjects themselves report (*a*) after four or five hours in isolation that they could not follow a connected train of thought, and (*b*) that their motivation for study or the like was seriously disturbed for 24 hours or more after coming out of isolation? The subjects were reasonably well adjusted, happy, and able to think coherently for the first four or five hours of the experiment; why, according to my theory, should this not continue, and why should the organization of behavior not be promptly restored with restoration of a normal environment?

You will forgive me perhaps if I do not dilate further on my own theoretical difficulties, paralleling those of others, but turn now to the conceptual nervous system of 1954 to ask what psychological values we may extract from it for the theory of motivation. I shall not attempt any clear answer for the difficulties we have considered—the data do not seem yet to justify clear answers—but certain conceptions can be formulated in sufficiently definite form to be a background for new research, and the physiological data contain suggestions that may allow me to retain what was of value in my earlier proposals while bringing them closer to ideas such as Harlow's (16) on one hand and to reinforcement theory on the other.

MOTIVATION AND C.N.S. IN 1954

For psychological purposes there are two major changes in recent ideas of nervous function. One concerns the single cell, the other an "arousal" system in the brain stem. The first I shall pass over briefly; it is very significant, but does not bear quite as directly upon our present problem. Its essence is that there are two kinds of activity in the nerve cell: the spike potential, or actual firing, and the dendritic potential, which has very different properties. There is now clear evidence (12) that the dendrite has a "slow-burning" activity which is not all-or-none, tends not to be transmitted, and lasts 15 to 30 milliseconds instead of the spike's one millisecond. It facilitates spike activity (23), but often occurs independently and may make up the greater part of the EEG record. It is still true that the brain is always active, but the activity is not always the transmitted kind that conduces to behavior. Finally, there is decisive evidence of primary inhibition in nerve function (25, 14) and of a true fatigue that may last for a matter of minutes instead of milliseconds (6, 9). These facts will have a great effect on the hypotheses of physiological psychology, and sooner or later on psychology in general.

Our more direct concern is with a development to which attention has already been drawn by Lindsley (24): the nonspecific or diffuse projection system of the brain stem, which was shown by Moruzzi and Magoun (34) to be an *arousal* system whose activity in effect makes organized cortical activity possible. Lindsley showed the relevance to the problem of emotion and motivation; what I shall attempt is to extend his treatment, giving more weight to cortical components in arousal. The point of view has also an evident relationship to Duffy's (13).

The arousal system can be thought of as representing a second major pathway by which all sensory excitations reach the cortex, as shown in the upper part of Fig. 1; but there is also feedback from the cortex and I shall urge that the *psychological* evidence further emphasizes the importance of this "downstream" effect.

In the classical conception of sensory function, input to the cortex was via

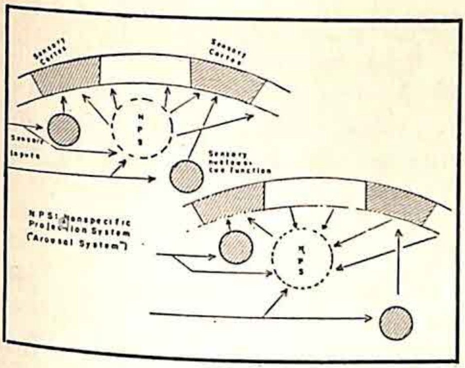

Sensory Cortex
Sensory Cortex
N
P
S
Sensory Inputs
Sensory nucleus: cue function
NPS: Nonspecific Projection System (Arousal System)
N
P
S

FIG. 1

the great projection systems only: from sensory nerve to sensory tract, thence to the corresponding sensory nucleus of the thalamus, and thence directly to one of the sensory projection areas of the cortex. These are still the direct sensory routes, the quick efficient transmitters of information. The second pathway is slow and inefficient; the excitation, as it were, trickles through a tangled thicket of fibers and synapses, there is a mixing up of messages, and the scrambled messages are delivered indiscriminately to wide cortical areas. In short, they are messages no longer. They serve, instead, to tone up the cortex, with a background supporting action that is completely necessary if the messages proper are to have their effect. Without the arousal system, the sensory impulses by the direct route reach the sensory cortex, but go no farther; the rest of the cortex is unaffected, and thus learned stimulus-response relations are lost. The waking center, which has long been known, is one part of this larger system; any extensive damage to it leaves a permanently inert, comatose animal.

Remember that in all this I am talking conceptual nervous system: making a working simplification, and abstracting for psychological purposes; and all these statements may need qualification, especially since research in this area is moving rapidly. There is reason to think, for example, that the arousal system may not be homogeneous, but may consist of a number of subsystems with distinctive functions (38). Olds and Milner's (37) study, reporting "reward" by direct intracranial stimulation, is not easy to fit into the notion of a single, homogeneous system. Sharpless' (40) results also raise doubt on this point, and it may reasonably be anticipated that arousal will eventually be found to vary qualitatively as well as quantitatively. But in general terms, psychologically, we can now distinguish two quite different effects of a sensory event. One is the *cue function,* guiding behavior; the other, less obvious but no less important, is the *arousal* or *vigilance function.* Without a foundation of arousal, the cue function cannot exist.

And now I propose to you that, whatever you wish to call it, arousal in this sense is synonymous with a general drive state, and the conception of drive therefore assumes anatomical and physiological identity. Let me remind you of what we discussed earlier: the drive is an energizer, but not a guide; an engine but not a steering gear. These are precisely the specifications of activity in the arousal system. Also, learning is dependent on drive, according to drive theory, and this too is applicable in general terms—no arousal, no learning; and efficient learning is possible only in the waking, alert, responsive animal, in which the level of arousal is high.

Thus I find myself obliged to reverse my earlier views and accept the drive conception, not merely on physiological grounds but also on the grounds of some of our current psychological studies. The conception is somewhat modified, but the modifications may not be entirely unacceptable to others.

Consider the relation of the effectiveness of cue function, actual or poten-

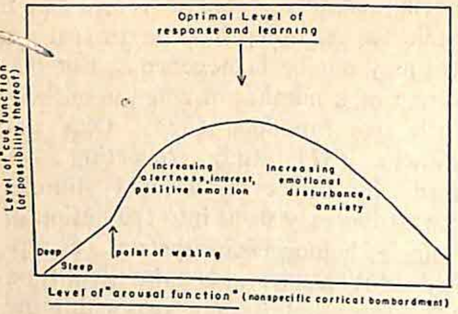


FIG. 2

tial, to the level of arousal (Fig. 2). Physiologically, we may assume that cortical synaptic function is facilitated by the diffuse bombardment of the arousal system. When this bombardment is at a low level an increase will tend to strengthen or maintain the concurrent cortical activity; when arousal or drive is at a low level, that is, a response that produces increased stimulation and greater arousal will tend to be repeated. This is represented by the rising curve at the left. But when arousal is at a high level, as at the right, the greater bombardment may interfere with the delicate adjustments involved in cue function, perhaps by facilitating irrelevant responses (a high D arouses conflicting $_sH_R$'s?). Thus there will be an optimal level of arousal for effective behavior, as Schlosberg (39) has suggested. Set aside such physiologizing completely, and we have a significant behavioral conception left, namely, that the same stimulation in mild degree may attract (by prolonging the pattern of response that leads to this stimulation) and in strong degree repel (by disrupting the pattern and facilitating conflicting or alternative responses).

The significance of this relation is in a phenomenon of the greatest importance for understanding motivation in higher animals. This is the *positive attraction of risk taking*, or mild fear, *and of problem solving*, or mild frustration, which was referred to earlier. Whiting and Mowrer (49) and Berlyne (4) have noted a relation between fear and curiosity—that is, a tendency to seek stimulation from fear-provoking objects, though at a safe distance. Woodworth (50) and Valentine (48) reported this in children, and Woodworth and Marquis (51) have recently emphasized again its importance in adults. There is no doubt that it exists. There is no doubt, either, that problem-solving situations have some attraction for the rat, more for Harlow's (16) monkeys, and far more for man. When you stop to think of it, it is nothing short of extraordinary what trouble people will go to in order to get into more trouble at the bridge table, or on the golf course; and the fascination of the murder story, or thriller, and the newspaper accounts of real-life adventure or tragedy, is no less extraordinary. This taste for excitement *must* not be forgotten when we are dealing with human motivation. It appears that, up to a certain point, threat and puzzle have positive motivating value, beyond that point negative value.

I know this leaves problems. It is not *any* mild threat, *any* form of problem, that is rewarding; we still have to work out the rules for this formulation. Also, I do not mean that there are not secondary rewards of social prestige for risk taking and problem solving—or even primary reward when such behavior is part of lovemaking. But the animal data show that it is not always a matter of extrinsic reward; risk and puzzle can be attractive in themselves, especially for higher animals such as man. If we can accept this, it will no longer be necessary to work out tortuous and improbable ways to explain why human beings work for money, why school children should learn without pain, why a human being in isolation should dislike doing nothing.

One other point before leaving Fig. 2: the low level of the curve to the right. You may be skeptical about such an extreme loss of adaptation, or disturbance of cue function and S-R relations, with high levels of arousal. Emotion is persistently regarded as energizing and organizing (which it certainly is at the lower end of the scale, up to the optimal level). But the "paralysis of terror" and related states do occur. As Brown and Jacobs (8, p. 753) have noted, "the presence of fear may act as an energizer . . . and yet lead in certain instances to an increase in immobility." Twice in the past eight months, while this address was being prepared, the Montreal newspapers reported the behavior of a human being who, suddenly finding himself in extreme danger but with time to escape, simply made no move whatever. One of the two was killed; the other was not, but only because a truck driver chose to wreck his truck and another car instead. Again, it is reported by Marshall (27), in a book that every student of human motivation should read carefully, that in the emotional pressure of battle no more than 15 to 25 per cent of men under attack even fire their rifles, let alone use them efficiently.

Tyhurst's (47) very significant study of behavior in emergency and disaster situations further documents the point. The adult who is told that his apartment house is on fire, or who is threatened by a flash flood, may or may not respond intelligently. In various situations, 12 to 25 per cent did so; an equal number show "states of confusion, paralyzing anxiety, inability to move out of bed, 'hysterical' crying or screaming, and so on." Three-quarters or more show a clear impairment of intelligent behavior, often with aimless and irrelevant movements, rather than (as one might expect) panic reactions. There seems no doubt: the curve at the right must come down to a low level.

Now back to our main problem: If we tentatively identify a general state of drive with degree of arousal, where does this leave hunger, pain, and sex drives? These may still be anatomically separable, as Stellar (45) has argued, but we might consider instead the possibility that there is just one general drive state that can be aroused in different ways. Stellar's argument does not seem fully convincing. There are certainly regions in the hypothalamus that control eating, for example; but is this a *motivating* mechanism? The very essence of such a conception is that the mechanism in question should energize *other* mechanisms, and Miller, Bailey, and Stevenson (31) have shown that the opposite is true.

But this issue should not be pressed too far, with our present knowledge. I have tried to avoid dogmatism in this presentation in the hope that we might try, for once, to see what we have in common in our views on motivation. One virtue of identifying arousal with drive is that it relates differing views (as well as bringing into the focus of attention data that may otherwise be neglected). The important thing is a clear distinction between cue function and arousal function, and the fact that at low levels an increase of drive intensity may be rewarding, whereas at high levels it is a decrease that rewards. Given this point of view and our assumptions about arousal mechanisms, we see that what Harlow has emphasized is the exteroceptively aroused, but still low-level, drive, with cue function of course directly provided for. In the concept of anxiety, Spence and Brown emphasize the higher-level drive state, especially where there is no guiding cue function that would enable the animal to escape threat. The feedback from cortical functioning makes intelligible Mowrer's (35) equating anxiety aroused by threat of pain, and anxiety aroused in some way by cognitive processes related to ideas of the self. Solomon and Wynne's (44) results with sympathectomy are also relevant, since we must not neglect the arousal effects of interoceptor activity; and so is clinical anxiety due to metabolic and nutritional disorders, as well as that due to some conflict of cognitive processes.

Obviously these are not explanations that are being discussed, but possible lines of future research; and there is one problem in particular that I would urge should not be forgotten. This is the cortical feedback to the arousal system, in physiological terms: or in psychological terms, the *immediate drive value of cognitive processes,* without intermediary. This is psychologically demonstrable, and *has* been demonstrated repeatedly.

Anyone who is going to talk about acquired drives, or secondary motivation, should first read an old paper by Valentine (48). He showed that with a young child you can easily condition fear of a caterpillar or a furry animal, but cannot condition fear of opera glasses, or a bottle; in other words, the fear of some objects, that seems to be learned, was there, latent, all the time. Miller (29) has noted this possibility but he does not seem to have regarded it very seriously, though he cited a confirmatory experiment by Bregman; for in the same passage he suggests that my own results with chimpanzee fears of certain objects, including strange people, may be dealt with by generalization. But this simply will not do, as Riesen and I noted (21). If you try to work this out, for the infant who is terrified on *first* contact with a stranger, an infant who has never shown such terror before, and who has always responded with eager affection to the only human beings he has made contact with up to this moment, you will find that this is a purely verbal solution.

Furthermore, as Valentine observed, you cannot postulate that the cause of such fear is simply the strange event, the thing that has never occurred before. For the chimpanzee reared in darkness, the first sight of a human being is of course a strange event, by definition; but fear of strangers does not occur until later, until the chimpanzee has had an opportunity to learn to recognize a few persons. The fear is not "innate" but depends on some sort of cognitive or cortical conflict of learned responses. This is clearest when the baby chimpanzee, who knows and welcomes attendant *A* and attendant *B*, is terrified when he sees *A* wearing *B*'s coat. The role of learning is inescapable in such a case.

The cognitive and learning element may be forgotten in other motivations, too. Even in the food drive, some sort of learning is fundamentally important: Ghent (15) has shown this, Sheffield and Campbell (41) seem in agreement, and so does the work of Miller and his associates (3, 32, 30) on the greater reinforcement value of food by mouth, compared to food by stomach tube. Beach (1) has shown the cortical-and-learning element in sex behavior. Melzack (28) has demonstrated recently that even pain responses involve learning. In Harlow's (16) results, of course, and Montgomery's (33), the cognitive element is obvious.

These cortical or cognitive components in motivation are clearest when we compare the behavior of higher and lower species. Application of a *genuine* comparative method is essential, in the field of motivation as well as of intellectual functions (22). Most disagreements between us have related to so-called "higher" motivations. But the evidence I have discussed today need not be handled in such a way as to maintain the illusion of a complete separation between our various approaches to the problem. It *is* an illusion, I am convinced; we still have many points of disagreement as to relative emphasis, and as to which of several alternative lines to explore first, but this does not imply fundamental and final opposition. As theorists, we have been steadily coming together in respect of ideational (or representative, or mediating, or cognitive) processes; I believe that the same thing can happen, and is happening, in the field of motivation.

REFERENCES

1. BEACH, F. A. The neural basis at innate behavior. III. Comparison of learning ability and instinctive behavior in the rat. *J. comp. Psychol.*, 1939, **28**, 225–262.
2. BERGMANN, G. Theoretical psychology. *Annu. Rev. Psychol.*, 1953, **4**, 435–458.
3. BERKUN, M. M., KESSEN, MARION L., & MILLER, N. E. Hunger-reducing effects of food by stomach fistula versus food by mouth measured by a consummatory response. *J. comp. physiol. Psychol.*, 1952, **45**, 550–554.
4. BERLYNE, D. E. Novelty and curiosity as determinants of exploratory behavior. *Brit. J. Psychol.*, 1950, **41**, 68–80.
5. BEXTON, W. H., HERON, W., & SCOTT, T. H. Effects of decreased variation in the sensory environment. *Canad. J. Psychol.*, 1954, **8**, 70–76.
6. BRINK, F. Excitation and conduction in the neuron. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 50–93.
7. BROWN, J. S. Problems presented by the concept of acquired drives. In *Current theory and research in motivation: a symposium.* Lincoln: Univer. of Nebraska Press, 1953. Pp. 1–21.
8. BROWN, J. S., & JACOBS, A. The role of fear in the motivation and acquisition of responses. *J. exp. Psychol.*, 1949, **39**, 747–759.
9. BURNS, B. D. The mechanism of afterbursts in cerebral cortex. *J. Physiol.*, 1955, **127**, 168–188.
10. BUTLER, R. A. Discrimination learning by rhesus monkeys to visual-exploration motivation. *J. comp. physiol. Psychol.*, 1953, **46**, 95–98.
11. CARPER, J. W., & POLLIARD, F. A. Comparison of the intake of glucose and saccharin solutions under conditions of caloric need. *Amer. J. Psychol.*, 1953, **66**, 479–482.
12. CLARE, M. H., & BISHOP, G. H. Properties of dendrites; apical dendrites of the cat cortex. *EEG clin. Neurophysiol.*, 1955, **7**, 85–98.
13. DUFFY, ELIZABETH. An explanation of "emotional" phenomena without the use of the concept "emotion." *J. gen. Psychol.*, 1941, **25**, 283–293.
14. ECCLES, J. C. *The neurophysiological basis of mind.* London: Oxford Univer. Press, 1953.
15. GHENT, LILA. The relation of experience to the development of hunger. *Canad. J. Psychol.*, 1951, **5**, 77–81.
16. HARLOW, H. F. Mice, monkeys, men, and motives. *Psychol. Rev.*, 1953, **60**, 23–32.
17. HARLOW, H. F., HARLOW, MARGARET K., & MEYER, D. R. Learning motivated by a manipulation drive. *J. exp. Psychol.*, 1950, **40**, 228–234.
18. HEBB, D. O. Elementary school methods. *Teach. Mag.* (Montreal), 1930, **12**, 23–26.
19. HEBB, D. O. *Organization of behavior.* New York: Wiley, 1949.
20. HEBB, D. O. On human thought. *Canad. J. Psychol.*, 1953, **7**, 99–110.
21. HEBB, D. O., & RIESEN, A. H. The genesis of irrational fears. *Bull. Canad. Psychol. Ass.*, 1943, **3**, 49–50.
22. HEBB, D. O., & THOMPSON, W. R. The social significance of animal studies. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge, Mass.: Addison-Wesley, 1954. Pp. 532–561.
23. LI, CHOH-LUH, & JASPER, H. Microelectrode studies of the cerebral cortex in the cat. *J. Physiol.*, 1953, **121**, 117–140.
24. LINDSLEY, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 473–516.
25. LLOYD, D. P. C. A direct central inhibitory action of dromically conducted impulses. *J. Neurophysiol.*, 1941, **4**, 184–190.
26. MACCORQUODALE, K., & MEEHL, P. E. A distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, **55**, 95–107.

27. Marshall, S. L. A. *Men against fire.* New York: Morrow, 1947.
28. Melzack, R. The effects of early experience on the emotional responses to pain. Unpublished doctor's dissertation, McGill Univer., 1954.
29. Miller, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 435–472.
30. Miller, N. E. Some studies of drive and drive reduction. Paper read at Amer. Psychol. Ass., Cleveland, September, 1953.
31. Miller, N. E., Bailey, C. J., & Stevenson, J. A. F. Decreased "hunger" but increased food intake from hypothalamic lesions. *Science,* 1950, **112**, 256–259.
32. Miller, N. E., & Kessen, Marion L. Reward effects of food via stomach fistula compared with those via mouth. *J. comp. physiol. Psychol.,* 1952, **45**, 555–564.
33. Montgomery, K. C. The effect of activity deprivation upon exploratory behavior. *J. comp. physiol. Psychol.,* 1953, **46**, 438–441.
34. Moruzzi, G., & Magoun, H. W. Brain stem reticular formation and activation of the EEG. *EEG clin. Neurophysiol.,* 1949, **1**, 455–473.
35. Mowrer, O. H. Motivation. *Annu. Rev. Psychol.,* 1952, **3**, 419–438.
36. Nissen, H. W. Instinct as seen by a psychologist. *Psychol. Rev.,* 1953, **60**, 291–294.
37. Olds, J., & Milner, P. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J. comp. physiol. Psychol.,* 1954, **47**, 419–427.
38. Olszewski, J. The cytoarchitecture of the human reticular formation. In E. D. Adrian, F. Bremer, & H. H. Jasper (Eds.), *Brain mechanisms and consciousness.* Oxford: Blackwell, 1954.
39. Schlosberg, H. Three dimensions of emotion. *Psychol. Rev.,* 1954, **61**, 81–88.
40. Sharpless, S. K. Role of the reticular formation in habituation. Unpublished doctor's dissertation, McGill Univer., 1954.
41. Sheffield, F. D., & Campbell, B. A. The role of experience in the "spontaneous" activity of hungry rats. *J. comp. physiol. Psychol.,* 1954, **47**, 97–100.
42. Sheffield, F. D., & Roby, T. B. Reward value of a non-nutritive sweet taste. *J. comp. physiol. Psychol.,* 1950, **43**, 471–481.
43. Sheffield, F. D., Wulff, J. J., & Backer, R. Reward value of copulation without sex drive reduction. *J. comp. physiol. Psychol.,* 1951, **44**, 3–8.
44. Solomon, R. L., & Wynne, L. C. Avoidance conditioning in normal dogs and in dogs deprived of normal autonomic functioning. *Amer. Psychologist,* 1950, **5**, 264. (Abstract)
45. Stellar, E. The physiology of motivation. *Psychol. Rev.,* 1954, **61**, 5–22.
46. Thompson, W. R., & Solomon, L. M. Spontaneous pattern discrimination in the rat. *J. comp. physiol. Psychol.,* 1954, **47**, 104–107.
47. Tyhurst, J. S. Individual reactions to community disaster: the natural history of psychiatric phenomena. *Amer. J. Psychiat.,* 1951, **107**, 764–769.
48. Valentine, C. W. The innate bases of fear. *J. genet. Psychol.,* 1930, **37**, 394–419.
49. Whiting, J. W. M., & Mowrer, O. H. Habit progression and regression—a laboratory study of some factors relevant to human socialization. *J. comp. Psychol.,* 1943, **36**, 229–253.
50. Woodworth, R. S. *Psychology.* New York: Holt, 1921.
51. Woodworth, R. S., & Marquis, D. G. *Psychology.* (5th Ed.) New York: Holt, 1947.

(Received December 14, 1954)

Psychological Review
Vol. 62, No. 4, 1955

PROBABILISTIC FUNCTIONING AND THE CLINICAL METHOD[1]

KENNETH R. HAMMOND
University of Colorado

It is probably true that most nonclinicians believe (with considerable justification) that the clinical method does not meet the criteria of science. This belief is ordinarily founded on the following grounds: The process by which the clinician arrives at a decision is private, quasi-rational, and nonrepeatable. Frequently, the clinician cannot report with confidence exactly how he arrives at a decision. He cannot point at the datum, or the configuration of data, which led to his decision. And, if he could, he would be doing nothing more than providing us with an introspective report. When two clinicians are involved, one introspection is left to stand against another. In brief, clinicians' judgments are a function of a process they cannot trace. It is as if we put our empirical data into a computing machine, the processes of which we did not understand and which frequently produced different results depending on which machine we used and when we used it.

These criticisms are difficult to meet; the effort to remove them has centered around the development of clinical tests. The aim of the test is to produce a retraceable process. However, this movement, while vigorous in a numerical sense—there are lots of tests—cannot quite ignore the skeleton in the closet. That is, in the last analysis, in the clinical situation the value of a test depends upon its agreement with *some* clinician's judgment. The objectivity of the test is no defense against the clinician's decision. All the test can do is to become a reasonable facsimile of the clinician. And since the clinician stands alone as the ultimate criterion, tests stand or fall through their agreement with a reduction process which remains a mystery.

Since the clinical decision is the ultimate criterion, the final measuring device against which other techniques are evaluated, and since this decision process is a *private* one, it is hardly surprising that there is some question as to whether knowledge is increasing in clinical psychology and psychiatry. Therefore, this paper proposes a change in point of departure.

The plan of the paper is this: First, two methodological issues are discussed—the partition between the observer and the object, and a distinction between two types of reduction bases. Second, these methodological issues are interpreted to fit with behavioral fact in general and the clinical situation in particular. Third, a method of research congruent with these behavioral facts is discussed. Finally, an example of the utility of this analysis is presented.

PARTITION BETWEEN SUBJECT AND OBJECT

Concerning the first of the methodological points, some remarks by Lenzen (6) in connection with physics are quite relevant. Although these remarks may or may not carry significance for psychologists above the level of analogy, they are presented here because they clarify a problem common to psychology and physics; i.e., interaction between observer and object.

---

[1] This paper was part of the Symposium on the Probability Approach in Psychology held at the Berkeley Conference for the Unity of Science, University of California, July, 1953. The author wishes to express his appreciation to Professor Egon Brunswik for suggesting his participation in the symposium.

According to Lenzen, the partition[2] between the object and observer shifts according to the intent of the observer. For example, "If a physicist is looking at a pointer on a scale, its status depends on the purpose of the observation. If he is using the instrument to measure an electric current, the pointer is an extension of the observer; the object is the electric current. If the physicist is calibrating his instrument, the pointer is part of the object of observation; the light by which the pointer is seen is then an instrument which belongs to the observer" (6, p. 29). Thus, the partition shifts according to the purpose of the investigation.

Now the same holds true for the clinical method. If the patient is being studied under usual circumstances, the partition stands between the subject being studied and the observer. Clearly the clinician is the observer, the subject or patient is the object being observed. If the clinician is studying a patient by means of a test, the test is an extension of the clinician just as a meter, say, is an extension of the physicist. If the clinician is studying a test, the scoring categories are the object of observation just as a pointer would be the object for a physicist calibrating a meter. But the clinician almost invariably stands beyond the partition as the observer and in most situations the partition is at the object—the patient.

There is an interesting parallel between difficulties in observation in the clinical situation and in the observation of microphysical entities. Lenzen remarks:

> In an observation of a micro-physical quantity there occurs an interaction between object and instrument; the instrument reacts against the object and may produce an unpredictable, finite change in the value of a quantity that is . . . being observed (6, p. 30).
>
> The observations of micro-physics require interpretation in terms of classical concepts but the fundamentally unpredictable, finite effects of the disturbances by the instruments of observation lead to a restriction in the applicability of classical concepts to micro-physical objects. The cognitive partition between object and apparatus is the seat of an indeterminacy which limits theoretical physics to the statistical prediction of the results of classically interpreted experiments (6, p. 31).

The parallel between this situation and the clinical situation is clear. The clinician certainly interacts with the object being observed, and, in principle, "may produce an unpredictable, finite change" in the object. Moreover, the object may produce an "unpredictable, finite change" in the observer.[3]

There are thus two points to be made here. First, in order to understand the interaction between the clinician-observer and patient-object, it is proposed that we shift the partition to a point

---

[2] Lenzen explicates the meaning of the term "partition" as follows (6, p. 28): "Tactual perception is an interaction between a body and end organs such as those in the tip of a finger. If one touches a desk with a finger, the partition is between them. An observer, however, may be extended by mechanical devices. Bohr has cited the following example: If one firmly grasps a long stick in one's hand and touches it to a body, the body touched is the object of observation, and the stick is an apparatus that may be viewed as part of the observer. It is a psychological fact that one locates the tactual aspect at the end of the stick, so that the partition is between the body and the end of the stick. If, however, the stick is held loosely in the hand, the stick becomes the perceived object, and the partition is between stick and hand."

[3] In a sense, the clinician when describing a patient gives a report on the changes which happened to himself. Note Lenzen's remark here: "In such observations (interaction between subject and object) it is not possible to control the action of the measuring instrument upon the object, for the instrument cannot be investigated while serving as a means of observation" (6). It appears likely that the clinical psychologist has so infrequently been the subject of investigation because he has been serving primarily "as a means of observation."

beyond the clinician-observer where our observations can take place in a noninteractive fashion. Second, it is proposed that we consider the "cognitive partition between object and [clinician]" to be an indeterminacy relation in a full theoretical sense, and utilize research procedures congruent with this proposition.

Stated otherwise, it is suggested that we consider the clinician not as a *reader* of instruments, as tradition has it, but as an instrument to be analyzed and understood in terms of a probability model. (An example of research following these propositions will be presented later.) We turn now to problems relating to a reduction base in the study of the clinical method.

INTERSUBJECTIVE COMMUNICABILITY

Consider next the problem of (macro-) physical measurement. The physical scientist begins with events which are *both* intersubjectively observable *and communicable*. That is, observers agree about a given event with a high degree of reliability and can readily communicate the reason for their response. Thus, observers of boiling and freezing points can agree not only as to when the liquid boils or freezes, but can point to, can communicate the basis for their decision.

Now consider the situation with regard to behavior. Observers of the state of anger may agree that such a state exists (i.e., high reliability may be achieved), *but* they may not be able to communicate the basis for their decision, or they may have decided on different evidence (i.e., intersubjective communicability is not achieved).

The crucial question here is this: Is noncommunicability[4] merely due to technical difficulties sooner or later to be surmounted? Or is noncommunicability a direct reflection of behavior, that is, a starting point for the study of behavior rather than a difficulty to be eliminated if possible? Should it actually be the case that noncommunicability is a direct reflection of behavior, then psychologists should cope with it theoretically rather than treat it merely as a technical difficulty. If, on the other hand, noncommunicability is merely due to poor circumstances of measurement, then we must turn to the laboratory where circumstances can be arranged more neatly. Here it will be hoped that communicability can be arranged via operational definition. If this can be done (and it is likely that most psychologists believe it can), then clinicians will have to wait for the technical problems of noncommunicability to be overcome in the laboratory.

It is also likely, however, that many psychologists have misgivings about the rapidity with which this goal is being reached. For it does appear that such a position accepts uncritically the hypothesis that noncommunicability is simply a function of inadequate apparatus, i.e., poor techniques. An equally tenable hypothesis would be that the apparatus is not inadequate, but rather that noncommunicability is a phenomenon to be understood rather than one to be eliminated from study. One would then be faced with the problem of developing a behavior theory and a research methodology appropriate to the problem. We now turn to the theoretical problem, that of considering observer-object interaction and noncommunicability jointly in relation to the theoretical concept of vicarious functioning.

VICARIOUS FUNCTIONING

The notion of noncommunicability can be robbed of its metaphysical air by consideration of the behavioral fact of *vicarious functioning*. Almost all students of behavior are in agreement with Tolman (8) and Brunswik (2) that higher organisms may substitute one form of behavior for another in order to achieve a goal. In the biological literature this phenomenon has been termed *equifinality* (2, p. 17). And concerning the perception of the environment, Brunswik and others have shown that cues to distance, say, may substitute for one another (1, p. 48). This phenomenon has been termed *equipotentiality* of cues.[5] Thus, vicarious functioning refers to the *variability* in what might be termed behavioral "output" (equifinality) and "input" (equipotentiality of cues for an organism).

---

[4] Perhaps a more nearly correct term here would be "limited intersubjective communicability." In the interests of simplicity the author prefers to risk overstating the case—thus, "noncommunicability."

Now the concepts of noncommunicability and observer-object interaction may be set in parallel to the concepts of equifinality and equipotentiality. Consider the clinical situation. The patient is trying, say, to achieve a certain goal. The clinician is attempting to discover the patient's motive. The patient substitutes one form of behavior for another as he attempts to achieve his goal (equifinality). The clinician perceives these behaviors, as they substitute for one another, as cues which also substitute for one another (equipotentiality). Because of vicarious functioning, then, the clinician is hard-pressed to point at, to communicate, the basis for a decision (except in the special case where univocal cues are available). Moreover, the partition between observer and object becomes indistinct for the same reason. *Vicarious functioning, then, lies at the heart of the private, quasi-rational nature of the clinical decision.* (Lenzen's phrase is certainly applicable here: "The cognitive partition between object and apparatus is the seat of an indeterminacy . . ." [6, p. 31].) Thus, assuming vicarious functioning (equifinality and equipotentiality) to occur, noncommunicability and observer-object interaction are not merely regrettable clinical occurrences to turn one's back on, but are starting points for the analysis of the clinical method, provided the appropriate research method is available.

REPRESENTATIVE DESIGN

What are the requirements of an appropriate method? An appropriate method must permit vicarious functioning to take place. It must take equipotentiality and equifinality as given, and it must permit inductive generalizations despite them.

Brunswik's development of a methodology which he terms "representative design" (1) seems to meet these criteria. Representative design is in part developed upon the concept of vicarious functioning, and requires that this fact of behavior not be eliminated (1, p. 23 f., 48 f.; 2, sec. 8).

The *uncritical* observance of the criteria of strict classical design (in contrast to representative design) of experiments is, to my mind, a principal stumbling block to the advance of clinical psychology. As long as vicarious functioning is ruled out of the experimental laboratory situation in accord with the tradition of nomothetic behaviorism, and as long as the partition between the clinical observer remains at the object, in accord with the tradition of clinical psychology, so long must clinical psychology and experimental psychology remain isolated disciplines. However, it might be hypothesized that noncommunicability and observer-object interaction are not merely technical difficulties but reflections of vicarious functioning. Accepting this hypothesis, and shifting the traditional place of the partition from between the clinician and object to a point *beyond* the clinician, will then permit clinician-patient interaction to be studied—provided the nomothetic ideal is relinquished and the principles of representative design are invoked.

---

[5] This matter is discussed at length by Brunswik (2, pp. 16–25).

Before turning to our examples it is worth noting some of Brunswik's remarks (2, p. 9) which will serve to emphasize the importance of the separation of subject and object as well as the ambiguities derived from vicarious functioning:

> Crucial turns in the history of ideas . . . are sometimes described as "Copernican revolutions." They define a succession of increasingly threatening blows to the pride of the ego; in psychoanalytic terms, the history of science is one of "retreating narcissism," or disentanglement of the objective from the subjective and wishful. Copernicus himself dethroned man's planet as the center of a faraway universe; Darwin dethroned the human species as the absolute master of the animal kingdom; Freud went still further and dethroned the conscious ego as the true representative of our own motivational dynamics. Kant and Gestalt psychology complete the picture by showing the subjectivity of the thing-language. Discovery of an ambiguous rather than univocal relationship between distant regions or variables seems to be at the root of most of such revolutions; the new "schools" protest the respective "constancy hypotheses."

This paper, then, protests the "constancy hypothesis" of the clinician, and the nomothetic bias of the laboratory psychologist.

We now turn to two examples of research which utilize features of representative design and which should serve to clarify the above remarks.

ILLUSTRATIONS

Although our suggestions above indicate that our research should begin with the clinician-patient situation, we have found it simpler to begin with the clinician-test situation. That is, instead of beginning with the clinician "measuring" or interpreting a patient via the interview situation, we found it easier to begin with the situation where the clinician "measures" or interprets the patient via a test. The principle is the same in both cases, however.

Our first example, then, concerns the situation where the clinician interprets the Rorschach test. (Any other interpretive psychological test could have been used, or an example could be taken from clinical medicine.) An investigation carried out by Todd (7) had as its purpose the study of the clinician as he perceives and responds to cues to the subject's intelligence provided by the Rorschach test. Further, Todd's intention was to carry out this study in the same manner as a perception psychologist working within the framework of representative design might study a subject's perception of size, i.e., when multiple cues to distance are available to the subject.

Think of the situation this way. Analogous to the meter stick for measuring the length of a body, we have a standard intelligence test. Analogous to physical cues for judging the length of the body we have Rorschach responses which can be categorized in various ways. The clinician's task is to estimate IQ from the Rorschach responses, or cues, just as the subject's task in the size constancy experiment is to estimate bodily size from physical cues in the environment.

First, we may ask, how well does the psychologist perform? Obviously, the answer depends on the factor of information, i.e., his level of performance depends on the kind and amount of information we give him.[6] For ten clinical psychologists judging the records of 78 patients, and provided with categorized responses alone, the median correlation coefficient between Rorschach-estimated IQ and Wechsler-Bellevue IQ was + .47. This is better than chance and fairly impressive. A reasonable

---

[6] The Rorschach test provides two general kinds of material: one consists of the entire protocol, which contains the subject's responses verbatim, and the other his categorized responses.

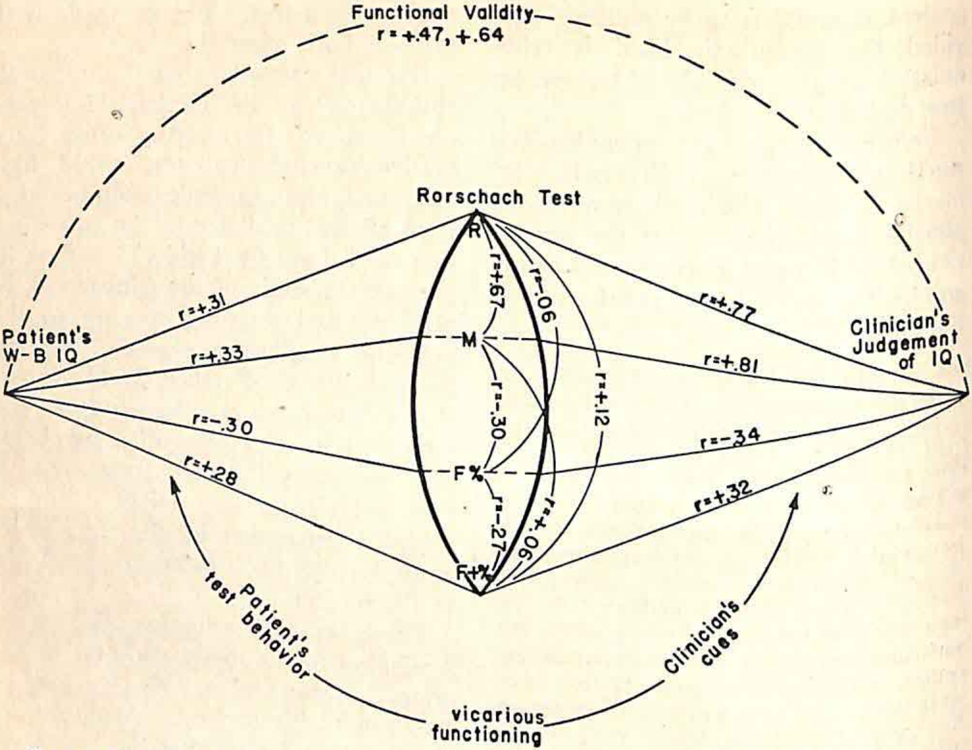


FIG. 1. Functional validity and mediating factors in clinicians' judgments of IQ from the Rorschach test.

question now is, would the addition of the verbal material in the subject's response help or hinder the clinician? Such addition does help. The correlation increases to + .64.[7]

Note that the first step follows our suggestion to set the subject-object partition which Lenzen speaks of at a point between the clinician and the *experimenter* in order that we may study the clinician. In the above example we discover the over-all functional validity of the clinician and Rorschach under two conditions—verbal material present and absent.

Todd next turns to the matter of vicarious functioning—the equifinality of responses and equipotentiality of cues. First is the matter of the relationship of various Rorschach factors to IQ. Results concerning this question are presented in the left half of Fig. 1. It is clear that Rorschach responses have a hierarchical probability relationship (indicated by the less than perfect correlations between Wechsler-Bellevue IQ and the four Rorschach factors) to Wechsler-Bellevue IQ for this population of subjects whose protocols are the stimulus objects for these clinicians.

The right half of Fig. 1 illustrates the relationship of certain cues to intelligence for the clinician. Here also we find a hierarchical probability relationship between Rorschach cues and clinicians' judgments of IQ.[8] (Intercue relationships between R and M, etc., are also shown in the figure.)

---

[7] Sophomores in elementary psychology, however, entirely naive with respect to the Rorschach, and estimating IQ on verbal characteristics alone, do practically as well, + .58. Apparently, therefore, training in the use of the Rorschach test adds little to proficiency, at least in this task.

[8] Cf. the data concerning perception and "cue-family hierarchies" presented by Brunswik, as well as his discussion of Hull's "habit-family hierarchy" (2).

These data are urged as empirical evidence for the concept of vicarious functioning in the clinical situation. They are also urged as evidence for the difficulties of communication under these circumstances.

It will be recalled that we suggested that the clinician be analyzed and understood in terms of a probability model. We now turn to this question: Will the multiple correlation procedure provide a good model for predicting the responses of the clinician?[9] That is, if a multiple regression formula is developed for each clinician, will it be possible to predict his responses (judgment of IQ) to a new set of Rorschach records? In an attempt to answer this question a multiple regression equation (based on the four factors best for each of the ten clinicians) was developed for each of ten clinicians on a sample of 39 Rorschachs, and predictions made for each clinician for a new sample of 39 Rorschach records. The responses predicted for the clinicians correlated quite highly with their actual responses. For the ten clinicians the median r was + .85, their individual correlations ranging from + .74 to + .92. Evidently the multiple correlation model which predicts that the clinician combines the data from the Rorschach in a linear, additive fashion is a good one—it predicts quite successfully in comparison with most psychological efforts.[10]

Neither Todd nor the author, however, would urge that the latter data be accepted as more than a surprising bit of evidence for the utility of the multiple correlation procedure as a theoretical model. The point is this: given the shift in locus of the partition between subject and object, applying the concept of vicarious functioning in a representative design framework does make it possible to set up *some* probability model to predict the indeterminacy relation suggested earlier.

Note that the design did not disturb the vicarious functioning of either responses or cues. Todd did not, for example, hold all Rorschach responses constant except one, or disrupt their intersubstitutability by arranging them orthogonally as in a systematic factorial design. It is precisely because representative design requires that vicarious functioning of cues be left undisturbed (and thus "representative") that we can specify the scope and precision of our induction for each clinician—such specification being based upon the nature of the sample of Rorschach protocols and the number of protocols in it. (Obviously inductions to *other clinicians* are foregone for sampling reasons.)

From a practical standpoint, it should be mentioned that certain variations were found among the ten clinicians in the effective use of various Rorschach factors. Certain clinicians were found to be using invalid cues, others neglecting valid ones. This kind of practical information should lead to higher predictive validities.

A study by Herring (4) illustrates a further application. Herring studied the problem of clinical psychologists' predictions of patients' responses to surgical anesthesia through the use of psychological tests. Because, according to medical authority, variability in response to surgical anesthesia can best be ascertained through medical clinical judgment, Herring utilized the conception of the clinical method outlined above on *both* sides of the experiment, the predictor side and criterion side.

---

[9] See Frenkel-Brunswik (**3**) for the first application of the multiple correlation approach to the analysis of trait-ratings. The reader will recognize the similarity of point of view taken here to that of Kelly and Fiske in reviewing the performance of clinical psychologists in the VA assessment project (**5**, pp. 200–202).

[10] As an aside, one might ask, as do Kelly and Fiske (**5**), if the clinicians are as efficient as a multiple correlation procedure? The median correlation between clinicians' judgments of IQ and Wechsler-Bellevue IQ was +.470 (see Fig. 1). A multiple R computed on the four most valid Rorschach factors was +.479. Evidently the clinicians are as *efficient* as a multiple correlation procedure. Those factors which were most valid also correlated most highly with the clinicians' judgments.

It may be seen from Fig. 1 that Todd worked only with the vicarious functioning of *psychological* variables. His "true" measure, IQ, was a normative one. Herring, on the other hand, dealt with a more common situation—no criterion available other than another expert's judgment. Ordinarily these two measures (judgments) are correlated, with no attention to the integration process involved on either side. Herring was able, however, through the application of the concept of both physiological vicarious functioning and psychological vicarious functioning, to trace through all these (partial) causal (?) chains so that the network of relationships became clear. That is, he was able to make some progress toward understanding the relationships between (and among) the cues to which the medical clinician responded, the cues to which the clinical psychologist responded, and, therefore *the reasons for correspondence and lack of correspondence between the medical and psychological clinicians' judgments,* i.e., the predictor and criterion variables. For example, Herring was able to show that one of three clinical psychologist's predictions did not correlate with the medical clinicians' criterion judgments because the psychologist's judgments were almost totally a function of one test—which happened to be invalid for this purpose.

SUMMARY

The attempt has been made here to scrutinize the clinical method from a systematic, methodological point of view. Lenzen's remarks concerning the partition between the subject and object were introduced in order to suggest that the clinician not be considered a reader of instruments, but an instrument to be understood in terms of a probability model. It was suggested that of two criteria for a reduction base, high reliability and communicability, the latter is difficult to achieve, not because of mere technical difficulties but because of a fundamental fact of behavior described as vicarious functioning. Brunswik's "representative design" is asserted to be the research procedure which is congruent with vicarious functioning; its applicability is demonstrated by Todd, who also demonstrated the feasibility of applying a probability model to the clinical situation.

REFERENCES

1. BRUNSWIK, E. *Systematic and representative design of psychological experiments.* Berkeley: Univer. of California Press, 1947. (Also in J. Neyman (Ed.), *Berkeley symposium on mathematical statistics and probability.* Berkeley: Univer. of California Press, 1949. Pp. 143–202.)
2. BRUNSWIK, E. *The conceptual framework of psychology.* Chicago: Univer. of Chicago Press, 1952. (*Int. Encycl. unified Sci.,* v. 1, no. 10.)
3. FRENKEL-BRUNSWIK, ELSE. Motivation and behavior. *Genet. Psychol. Monogr.,* 1942, 26, 121–265.
4. HERRING, F. H. A psychological study of patient response during surgery and anesthesia. Unpublished doctor's dissertation, Univer. of Colorado, 1954.
5. KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology.* Ann Arbor: Univer. of Michigan Press, 1951.
6. LENZEN, V. F. Procedures of empirical science. Chicago: Univer. of Chicago Press, 1938. (*Int. Encycl. unified Sci.,* v. 1, no. 5.)
7. TODD, F. J. A methodological study of clinical judgment. Unpublished doctor's dissertation, Univer. of Colorado, 1954.
8. TOLMAN, E. C. Cognitive maps in men and rats. *Psychol. Rev.,* 1948, 55, 189–207.

(Received June 11, 1954)

Psychological Review
Vol. 62, No. 4, 1955

FEEDBACK THEORY AND THE REFLEX ARC CONCEPT

CHARLES W. SLACK
Princeton University

In 1896 John Dewey criticized the new reflex arc concept in psychology on a number of grounds and concluded that: ". . . the distinction of sensation and movement as stimulus and response respectively is not a distinction which can be regarded as descriptive of anything which holds of psychical events or existences as such" (1, pp. 369–70). Dewey used the familiar example of the child reaching for the flame, taken from William James. The theory he severely criticized was the one which simply stated that the sensation of light is a stimulus to the grasping as a response, etc.

It was not until nearly a half-century later that experiments were performed which would allow a testing of the reflex arc concept in gross behavior—eye-hand coordination situations where some continuous record of the position of the "child's hand" and of the "flame" could be made. The experiments which allow for this kind of recording are called tracking experiments. The "flame" is now called the *target;* the "finger tip" becomes a special example of a number of *controls*, including variously loaded joy sticks and handwheels; the field of view, including that which the child sees of his hand, of the candle and of various reference objects, is now called the *display*.

Except for the fact that *S* is now usually limited to one or two dimensions in which he can move, and that his motivation is directed by instruction rather than by curiosity or burning pain, everything is pretty much the same with today's "human operator" and yesterday's *enfant terrible*. At least the classic example of the reflex arc is no better as an example than is the modern one.

There are but two things left to do to complete the similarity between the two examples. We must restrict the modern tracking to step-function inputs —sudden displacements of the target, in order to make it correspond to the sudden observance (or sudden lighting, if you will) of the candle in the classic example. Further, we shall only concern ourselves with the first half of the Dewey-James example where the child reaches for the flame. This phase illustrates the negative feedback principle to be discussed in this paper.

Tracking of step-function inputs, then, is a good example of the reflex arc. It is also a very special case of the psychophysical method of reproduction where the error contributed by *S* may eventually be reduced to zero. That is, tracking is the method of reproduction with knowledge of results and usually with continuous recording over time of the target and the control, or at least the difference between them.

Today's tracking, however, usually tends to imply more than a particular experimental situation. There is a theory which goes along with most experiments of this sort: on the one hand it is a more precise statement of the reflex arc notion, and on the other hand it says some quite different things.

Briefly stated, the theory of negative feedback extends to all systems which use a measure of the difference between the target and the control to decrease that difference (3). In the Dewey-James example, the difference between the finger tip and the flame is fed back and used to control perhaps the velocity of the hand; the greater the distance to go, the faster the movement.

We shall not concern ourselves here with the specific technical advantages of particular methods of analysis of feedback systems or with the special assumptions which many of these methods employ, but will turn instead to the general notion of the feedback system as a model for the reflex arc situation. The purpose of this paper is: first, to ascertain whether or not feedback theory (or servo theory) answers any of the criticisms of reflex arc theory which were raised by Dewey in 1896; and second, to see in what ways feedback theory conflicts with the simple stimulus-response notion implied in the reflex arc theory of Dewey's day.

First of all we notice that by its very nature of being a closed-loop theory, or a theory applying to systems which continually or continuously use a measure of performance to control performance, feedback theory meets Dewey's objection that "[reflex arc] gives us literally an arc, instead of the circuit; and not giving us the circuit of which it is an arc, does not enable us to place, to center, the arc" (1, p. 370). Reflex arc theory, as criticized by Dewey, was apparently an open-loop system. Servo theory closes the loop.

Secondly, Dewey criticized the reflex arc notion as follows:

> Upon analysis, we find that we begin not with a sensory stimulus, but with a sensorimotor coordination, the optical-ocular, . . .
>
> Now if this act, the seeing, stimulates another act, the reaching, it is because both of these acts fall within a larger coordination; . . . the ability of the hand to do its work will depend, either directly or indirectly upon its control, as well as its stimulation, by the act of vision . . . The reaching, in turn, must both stimulate and control the seeing . . . (1, p. 358).

Since the analysis of a feedback system is in no way vitiated if the system is made up of a number of subsystems, it would seem that modern servo theory again fills the bill.

Thirdly, Dewey stated:

> The sensation or conscious stimulus is not a thing or existence by itself; it is that phase of a coordination requiring attention because, by reason of the conflict within the coordination it is uncertain how to complete it. . . . The end to follow is, in this sense, the stimulus. . . . From this point of view the discovery of the stimulus is the "response" to possible movement as "stimulus." . . . Generalized, sensation as stimulus, is always that phase of activity requiring to be defined in order that a coordination may be completed (1, p. 368).

This criticism of the reflex arc strikes at the heart of the matter, for it challenges the validity of the stimulus and response as separate entities which are causally related in some way.

We may expand upon this argument by means of a simple example from tracking:

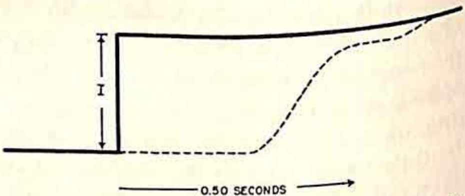


Fig. 1. Diagram of typical step-function input and output. The solid line is the input presented to *S* through a narrow slit at a fast rate. The dotted line represents *S*'s attempt to keep his control pencil on the input. The error in initial response may be an undershoot (as shown), an overshoot, or zero.

Figure 1 shows a diagram of a typical step-function input (solid line) and output (dotted line). What was presented to *S* was a dot remaining motionless at one position on the field, and then its instantaneous displacement (*I*) to another position on the field. The dotted line shows *S*'s attempt to follow this displacement of the target with his own control. His control in this case was a pencil which he attempted to keep on a short segment of line (the solid line), which he viewed through a narrow slit.

His output was then recorded, superimposed upon the input (which *E* programmed for him by drawing lines at various positions and of various lengths, corresponding to time intervals, on a strip of paper which was fed past the slit where *S*'s pencil point was resting).

This tracking input-output diagram might be taken as a slightly simpler example of the reflex arc situation than the child-candle.

In order better to understand Dewey's objection to the naive stimulus-response description of this type of behavior, we might attempt to apply stimulus response notation to the input-output diagram of tracking. Let us define the *stimulus* as a change in environment over time produced and measured by *E* by operations which he describes in such a way that they may be duplicated by others, and: (*a*) which, at any time, might provide some or all of the information leading to the response at that time, or (*b*) which, at any time, might possibly determine all or part of the response at that time, or (*c*) to which, at any time, the response might be related in some way, or (*d*) which is a necessary condition for the response.

Now let us check through the characteristics of the input to find what meets the criteria listed above.

First, the mere *existence of the target dot* somewhere (anywhere) on the field cannot alone be called stimulus. The existence of the target dot is not a *change* in environment over time; that is, the dot may always be there during the experiment. Furthermore, it meets none of the other criteria. The existence of the target dot might be considered a necessary condition for the response were it not for the fact that under certain conditions (after the establishment of expectancies of various kinds) the response is obtained without the existence of the target dot (5). At any rate the existence of the target provides no information for the response, cannot possibly determine the response, and the response is not much related to the mere existence of the target.

Second, the *position of the dot relative to the rest of the field* (slit) is clearly not the stimulus. In the first place, we discover that the same response may be obtained to targets positioned anywhere on the field. In the second place, the position of the dot on the field can in no way determine the extent of movement of *S*, nor provide information to determine the extent of movement. The *S* has to know "which way" and "how far" before he moves, and the position of the dot relative to the field does not tell him this. The position of the dot on the field (or indeed, the position of the dot relative to any frame of reference of *E*'s) is not related to the response of *S* within a wide range of conditions.

Third, the *displacement of the target* (the size of the movement *I*) cannot be called the stimulus since, if it were, we would be restricted only to the condition shown where *S*'s position is equivalent to the position of the target just as displacement occurs. The displacement of the target, like the position of the target, does not tell *S* "which way" or "how far" except for the time right after the target moves and just before *S* moves. After *S* has started to move (either in the right or the wrong direction), *I* can only be misleading information telling him "how far" or "which way."

The result of all this rather simple reasoning (not to be confused with Dewey's more eloquent generalizations) is that we come to the conclusion that there is nothing about what *E* does to the environment which can alone be called the stimulus if we give the term stimulus any psychological relevance.

Let us take a more functional view of the situation in an attempt to discover what might be the "stimulus" as far as *S* is concerned.

What must *S* know in order to be able to respond adequately? We gave the answer above when we said that he must know "which way" and "how far." In order to know these, he must know "where he is" and where he "wants to be." But the difference between where he is and where he wants to be, including sign, is the answer to the questions "which way" and "how far." To the extent, then, that we can assume that *S*'s frame of reference relative to which he knows these things corresponds in important ways to the one *E* gave him (the slit), and furthermore to the degree that where *S* wants to be corresponds to where *E* wants him to be, we can say that the stimulus is equal to the "just past" difference between the dotted line and the solid line. Or, to paraphrase the statement made earlier, the difference between where *S* is and where he wants to be is continually used as the "stimulus" to decrease that difference. If we know where *S* is (in the same regard as he knows it) and if we can assume that where *S* wants to be is equivalent to the position of the target relative to that regard, then, and only then can we specify what the "stimulus" is. But notice this about the "stimulus" which we have thus deduced. This "feedback" stimulus is much closer to Dewey's idea of the stimulus than it is to the stimulus response notion which demands that the stimulus be defined a priori in terms of *E*'s frame of reference. The Dewey feedback stimulus is far from a given, constant thing. It is an assumption—both on *E*'s part and upon *S*'s part. It is a continually changing thing. It depends upon where *S* is and *S* is moving. The *S* alters (and at times creates) the stimulus in just as real a sense as does *E*. Furthermore, the very existence of the stimulus depends upon a "common ground" or frame of reference between *E* and *S* and upon common purposes between them. Both the child and the man who lights the candle are parties to a transaction which has no meaning in terms of either one separately.

The "real" stimulus in the reflex arc is the error signal (difference between desired state and present state) and it is not the input defined in terms of *E*'s frame of reference.

This statement is all right as far as it goes but it does not go far enough. The error signal, as *E* knows it, *is* the difference between the dotted line and the solid line. But *S* has at his command, of course, *only an estimate* of this difference. There are threshold properties to be considered and there are constant and variable errors. Furthermore, there is much good evidence to show that *S*'s estimate of how far and which way he has to go is influenced to some extent by his past experience: there are serial order effects and "range" effects (2). Perhaps other more complex dependencies may affect *S*'s estimate, especially at high speeds of performance (6).[1] Are we to include all this in the term stimulus? If we are to give the term *any* a priori psychological relevance we must consider *some* characteristics of *S* in our definition. If we do not want to be arbitrary about it, we are forced to qualify our definition of the term "stimulus" with the following clauses:

1. It is an abstraction out of a process, called, by Dewey, a transaction.
2. It is created by *S* as well as by *E*.
3. It is never fixed or constant unless both *E* and *S* want it to be.
4. It is an assumption.
5. It is at least a difference between two estimates—one with regard to the

---

[1] The difference which exists between compensatory and follow tracking indicates that *S* is using more than the error signal upon which to base his response.

control and one with regard to the target.

6. Experience and various psychological errors enter into this estimate.

We may avoid the above argument by defining stimulus purely in physical terms, that is, relative to *E*'s frame of reference. We may define it as the existence, or the position, or the displacement, or the brightness of the target or as all of these. But if we do this—i.e., define stimulus as independent of *S*, it is going to have *no* a priori psychological relevance.

For purposes of clarity, we should have a word to stand for the heavy line in Fig. 1. We shall call this the input (input stimulus if you like), understanding that we can never be sure what, if anything, it has been put *into*. The dotted line in Fig. 1 we call the output. When we use these words, we should remember that these are not psychological variables. Nor are they physical variables which are usually relevant to a psychological understanding of this transaction. The input is defined as what *E* does to the target, relative to the frame of reference he chooses. This is to give input an operational definition relative to his purposes as an experimenter. The output, while it can be defined relative to *E*'s frame of reference, can only have existence for *S* relative to *S*'s purposes. Only in the event that *S*'s and *E*'s purposes overlap, that they have some goals in common, is an experiment possible.

Now, what is true for the "real" stimulus (as opposed to the "irrelevant" stimulus described above)—namely, that defining it depends upon common purposes between *E* and *S*, upon common frames of reference, that it can be created and altered by both *E* and *S*, and that it is an abstraction out of a complex transaction—is likewise true for the response.

CONCLUSION

We have taken one of the simplest examples of the reflex arc and attempted to apply stimulus-response concepts to it. In trying to do this we have discovered that we do not know what the stimulus is unless we know what the response is and what previous stimuli and responses were and that, as a matter of fact, we need quite a good understanding of the transaction in order usefully to call anything the stimulus. The same line of argument can be used for the so-called response side. We do not know what the response is unless we know what the stimulus is. Feedback theory tells us that the stimulus is at least the difference between the desired state and the present state. Our knowledge of psychology tells us that in this sense it must be a good deal more.

REFERENCES

1. DEWEY, J. The reflex arc concept in psychology. *Psychol. Rev.*, 1896, 3, 357–370.
2. ELLSON, D. G., & WHEELER, L., JR. The range effect. Dayton, O.: USAF, Air Materiel Command, Wright-Patterson Air Force Base, 1949. (AF Tech. Report No. 5813.)
3. FITTS, P. M. Engineering psychology and equipment design. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1287–1340.
4. KILPATRICK, F. P. (Ed.) *Human behavior from the transactional point of view.* Hanover: Institute for Associated Research, 1952.
5. SLACK, C. W. Learning in simple one-dimensional tracking. *Amer. J. Psychol.*, 1953, 66, 33–44.
6. SLACK, C. W. Some characteristics of the "range effect." *J. exp. Psychol.*, 1953, 46, 76–80.

(Received June 4, 1954)

Psychological Review
Vol. 62, No. 4, 1955

INTERPERSONAL BEHAVIOR AS INFLUENCED BY ACCURACY OF SOCIAL PERCEPTION

IVAN D. STEINER
University of Illinois

This paper presents a critical examination of two propositions which link accurate social perception with competence in interpersonal behavior and with group efficiency. The first of these propositions maintains that the more knowledge an individual has concerning the intentions, preferences, and beliefs of other persons, the more effectively he can participate in group activity with those other persons. This proposition provides the rationale for much of the training we give to teachers, social workers, clinical psychologists, and others whose work involves continuing interaction with people.

Attempts to verify this proposition have ordinarily involved the use of high sociometric status or high sociometrically evaluated productivity as indexes of the individual's interpersonal competence. With these indexes as criteria, considerable empirical evidence has been produced in support of this proposition. Chowdhry and Newcomb (3) found that persons who were sociometrically chosen, by various criteria, tended to make the most accurate estimates of group opinion on topics relevant to the group's aims and functions. Gage (8) found that high school seniors who did the most accurate job of predicting the responses which others would make to the Kuder Preference Record also received a large percentage of their classmates' sociometric choices. Greer, Galanter, and Nordlie (10) have reported that sociometrically popular members of infantry squads are more accurate than less popular members in predicting the sociometric positions of men in their squads. Norman (17) analyzed data obtained by a Veterans Administration research project and found that graduate students who were most often rejected by their classmates had the least realistic perceptions of others. Other investigators (2, 4, 9, 21) have also reported positive relationships between accuracy of social perception and interpersonal competence.

The second proposition to be examined here involves an extension of the first. It maintains that groups composed of individuals with accurate social perceptions will be more efficient than groups composed of members with less accurate social perceptions. Faith in this proposition guides much of our marriage counseling and has sometimes been a factor in the selection of work crews and play groups. Much of the empirical evidence in support of this proposition has been provided by studies which were not directly concerned with group efficiency. Thus, Festinger and his associates (6) have frequently given group members erroneous impressions of one another and have found that such impressions can lead to restricted communication within the group, rejection of members, and to lowered group cohesiveness. Such findings strongly suggest that inaccurate social perception can reduce the efficiency with which groups achieve either individual or group goals. Similar indirect support for the proposition is provided in a study by Dymond (5). She found that members of happily married couples (presumably groups which are efficient in achieving individual and group goals) made fewer errors than did members of unhappily married couples in predicting one another's responses to the Minnesota Multiphasic Personality Inventory. More direct support is provided by Cottrell and Dymond (4), whose ratings of group efficiency were positively correlated with the average accuracy of group members in predicting one another's self-ratings. Greer, Galanter, and Nordlie (10) have also provided direct support for the proposition. They found that infantry squads consisting of men who did a good job predicting one another's sociometric positions within the group were rated more highly than other squads on performance of field maneuvers.

Although the studies cited above provide some basis for confidence in each of the two propositions, other studies have failed to produce positive findings. Thus, the first proposition appears to be contradicted by Campbell's (18) failure to obtain a positive relationship between the accuracy with which naval officers judged the attitudes of their men and the officers' popularity with the crew. Similarly, Sprunger (20) failed to obtain a positive correlation between the accuracy with which office supervisors perceived their subordinates and the popularity of those supervisors with their workers. Perhaps the most notable failure to confirm the second proposition is contained in the work of Fiedler, who has studied the perceptual accuracy of members of efficient and inefficient basketball and surveying teams. His findings with respect to the effect of perceptual accuracy have been generally negative.[1] Indeed, there has sometimes appeared to be a slight association of perceptual accuracy with ineffective performance.

The existence of contradictory findings suggests that the two propositions are neither completely true nor completely false. It is probable that each is a true statement of a relationship which exists under certain limiting conditions, but a false statement when those limiting conditions do not prevail. If this is the case, it is desirable that the limiting conditions be identified and that the propositions be altered to take them into account.

Such limiting conditions might be identified by either of two methods. One method calls for a systematic and comprehensive analysis of the studies which have produced positive findings, and a comparison of those studies with others which have produced negative findings. Although in principle this method should be capable of identifying some of the conditions which determine the validity of the propositions, it is a procedure which involves all of the practical difficulties encountered in any ex post facto analysis. The other method seeks to identify the limiting conditions through an examination of the assumptions which are implicit in the two propositions. These implicit assumptions may themselves specify some of the conditions under which the two propositions are likely to be valid.

This paper will employ the second method in an attempt to identify certain of the conditions which must exist in order that accurate social perception can lead to increased interpersonal competence and group efficiency.

ASSUMPTIONS CONCERNING COLLECTIVE ACTION

The term "collective action" will be used in this paper to designate the behaviors of two or more persons who are attempting to satisfy needs or attain goals through copoerative effort in a "face-to-face" situation. Certain assumptions concerning the nature of collective action are implicit in the propositions cited above. It is implied that efficient collective action is accomplished through a dyadic process in which each participant adjusts his own behavior in response to the intentions and preferences of others. This view of collective action has received wide support in the literature. Mead (14) has contended that integrated collective action requires that each individual "take the role of the other," a process by which he gains insight into others' intentions and preferences. Asch (1) has eloquently defended a theory of human interaction in which the existence of mutually shared psychological fields and the capacity for "taking into account the emotions and thoughts of others" provide the basis for coordinated group behavior. In a recent paper Newcomb (16) has pictured the communication process as one in which two or more participating individuals seek "simultaneous orientation" to one another and to the events round about them. Such simultaneous orientation may be described as a relationship in which each individual takes cognizance of, and adjusts to, the attitudes of others. Newcomb says this relationship is essential to human life. Sears (19) has contended that "a dyadic unit can be derived from the combining of two or more monadic units" by taking into account the expectancies which one individual has of another's behavior. It is these expectancies which are "responsible for maintaining the stability of the dyadic unit." Hebb and Thompson (11) have proposed a similar explanation of collective action. Because of man's ability to anticipate not only his own next act and its effect, but also those of others, humans are capable of engaging in a kind of teamwork which is definitely lacking among other mammals.

---

[1] Fiedler, F. E. Personal communication, 1954. Fiedler (7) has recently summarized other findings of these studies.

It is easy to agree with these writers; efficient collective action often does require that each participant anticipate the behavior of others. If individuals engage in role taking, maintain simultaneous orientations, or have mutually shared fields, their collectively produced behavior sequences may constitute what Sears (19) has called a "dyadic system." Group behavior can become "more than a fortuitously useful conjunction of individual actions" or a mere piling up of "parallel monadic sequences." It can be interlocking behavior which leads to the achievement of individual or group goals with a minimum of wasted effort or needless duplication of behavior. However, there are reasons for doubting that this conception of collective action is universally valid.

Basic to this conception of collective action is the assumption that individuals are relatively free agents who are permitted to alter their own behaviors in response to their perceptions of the intentions or preferences of others. Furthermore, it is assumed that the changes which such perceptions produce in the individual's behaviors will lead to a more thoroughly integrated dyadic system. Unless these assumptions are met, accurate social perception, whether it be achieved through role taking, simultaneous orientation, or shared psychological fields, can scarcely be expected to increase interpersonal competence or group efficiency. In the remainder of this paper it will be contended that much of our most efficient collective action occurs when neither of these assumptions is met, and under circumstances which make accurate social perception relatively unimportant.

Even when these two assumptions are met, accurate social perception may fail to permit or encourage efficient collective action. It should be expected to do so only when the attributes, preferences, or intentions which are accurately perceived are relevant to the task at hand. Accurate perception of irrelevant qualities in the other person may only prove to be distracting in its effect. In the remainder of this paper "accurate social perception" will refer to the perception of qualities, preferences, and intentions which are relevant to the activities of the group.

ROLE SYSTEMS AND SOCIAL PERCEPTION

Role systems often permit or compel individuals to produce efficient dyadic behavior sequences even though they know little about one another's intentions or preferences. This contention will be clarified by the construction of a model with which to represent some of the characteristics of a role system.

It will be helpful to imagine a matrix of columns and rows, creating a system of cells. Each of the columns will bear the name of a category of persons who are recognized as functionally distinct by the members of the collectivity. Such categories may fall along any of a number of dimensions like age, sex, material wealth, occupation, etc. It is not at all necessary that the categories be unidimensional, and in the case of most collectivities they will not be. It is only necessary that the members of the collectivity be capable of distinguishing between the people who belong in different categories. This requirement is easily met, for if people are not naturally distinguishable they may be made artificially distinguishable through the use of special clothing, insignia, spatial arrangements, and a variety of other means.

The rows in the matrix will each carry the name of a category of situations. Although groups may seldom categorize situations as explicitly as they categorize persons, the multiplicity of stimulus situations precludes the possibility that each can be responded to as something totally unique. As Merton (15) has pointed out, groups tend to evolve standard ways of classifying the situations which are most frequently encountered in collective action. Similar situations tend to be classified into the same category and to be treated as though they were equivalent. Here, as in the case of categories of persons, there need be no apparent logic to the system. It is only necessary that participants be capable of assigning each situation to its proper niche. Standard assignment of situations to their appropriate categories can be achieved through a variety of socialization and indoctrination procedures.

Into the cells which are created by the intersection of columns and rows will be placed the designation of the special behavior which is enforced by the collectivity. Thus, if the collectivity is a symphony orchestra, the cell created by the intersection of the column "first violinist" and the row "third bar, Brahms' Lullaby" might carry instructions to produce specific musical tones at a specific tempo. Not all cells will receive unique entries, and some cells may receive no entry at all because no special behavior is required of a category of persons in the situation represented by a row in the matrix.

In terms of this model, a role consists of all those behaviors which are in the cells of a single column of a complete matrix. In the case of a symphony orchestra, the role of each participant is ordinarily described in detail by the musical score and each musician will know his own role though he may have only a limited understanding of other persons' roles. The role system is represented by the entire matrix and includes the role of each participant. It constitutes a system because the various roles are interlocked in a way which produces highly integrated collective action. The behaviors expected of one person supplement those expected of others. Often, as in the case of a work group, the role of one participant cannot be enacted unless others are enacting their roles properly, and the inadequate performance of any one role may disrupt the entire system.

It is the advantage of role systems that behavior synthesis has been incorporated into the system itself, and that participants need not infer the strategies of their associates or improvise an effective synthesis as they engage in collective action. They need only to produce the behaviors which are situationally appropriate for members of their categories. To be sure, other people and their behaviors will often be important elements in the situations to which the individual must respond. But this ordinarily means only that each individual must take into account the overt behavior of other categories of persons; it does not mean that he must make inferences concerning the idiosyncratic intentions or preferences of *unique* individuals. Although participants may develop enduring expectations concerning the unique qualities of specific other individuals, these expectations are not part of the role system. Indeed, participants in a role system are often required to ignore unique qualities and to treat one another as though each were a standard representative of a whole category. The individual who accurately projects himself into the unique psychological fields of others is likely to find it difficult to perform as a standard person or to treat others as such. And as Merton (15) has observed, any deviation from standard performance or standard treatment of others is likely to disrupt the whole role system. Consequently, those individuals who deviate from standard practice are likely to be ostracized. Given the existence of a role system, perceptual accuracy which goes beyond that required for correct categorization of others may decrease group efficiency and reduce interpersonal competence. In this connection it is instructive to recall that Hemphill (12) found that members of large groups (30 or more persons) tended to prefer leaders who did not treat them as unique individuals.

A large proportion of all collective action must be presumed to occur within the framework of role systems where the intentions and preferences of other persons may be largely irrelevant and where the individual's freedom to adjust to such intentions and preferences is, in any case, highly restricted. Naval crews, office staffs, and surveying teams are probable examples of groups which do have elaborate role systems. It is to be noted that investigations of such groups have generally failed to obtain a positive relationship between accuracy of social perception and group efficiency or interpersonal competence. A possible contradiction of this trend is represented by the study of infantry squads (10) in which accuracy in predicting the sociometric status of squad members correlated with group efficiency. However, accuracy in perceiving a status hierarchy is a particular type of accuracy which may be a by-product of efficient use of a role system. If, as may be expected, the efficiency of a squad is a function of the degree to which its members adhere to a role system, then efficient squads will be those which demand that members enact specified roles, and deviation from standard role performance will serve as a major reason for rejection of members. Members of efficient squads would then tend to make accurate predictions of one another's positions in the status hierarchy because each will be aware of the major criterion which has guided the creation of that hierarchy. In squads where adherence to the role system is less highly stressed, efficiency should be lower and members should have more difficulty predicting the number of sociometric choices each man will receive. Until the correctness of this explanation is ascertained, it is probably unwise to regard this study as a clear contradiction of what appears to be a general trend.

Although the published reports of research in this area are not sufficiently detailed to permit categorical statements, the reported findings seem to support the contention that efficiency in groups which operate within the framework of a role system is not increased by accurate understanding of others' preferences and intentions. What are needed here are controlled studies in which groups with and without the benefit of a role system are confronted by situations requiring coordinated action. If role systems do, in fact, have the properties described above, accuracy of interpersonal understanding should have no relationship (or perhaps a negative relationship) to the efficiency of groups which have appropriate role systems. On the other hand, there should be a positive relationship between the perceptual accuracy and the efficiency of groups without role systems, assuming, of course, that members are motivated to cooperate.[2]

SUMMARY

This paper has examined two propositions which maintain that accurate social perception is responsible for increased interpersonal competence and group efficiency. Although such propositions have provided a basis for considerable "applied psychology," empirical research has sometimes raised doubts concerning their validity. In order to discover some of the limiting conditions under which the propositions are likely to be valid, it has been necessary to examine the conception of collective action on which they are based.

This examination has led to the following conclusions. Accurate social perception should promote interpersonal competence and group efficiency if: (*a*) the group members are motivated to cooperate; (*b*) the accurately perceived qualities are relevant to the activities of the group; (*c*) members are free to alter their own behaviors in response to their perceptions of other members; and (*d*) the behavioral changes which are a consequence of accurate social perception are the kinds which produce a more thoroughly integrated dyadic system. Whenever any *one* or more of these conditions is not met, accurate social perception should fail to have the effect predicted by the two propositions which have been the major concern of this paper.

It has been contended that much of our most highly integrated and efficient collective action must be presumed to occur within the framework provided by role systems. Whenever this is the case, individual participants are neither required nor permitted to let their perceptions of other people's intentions or preferences affect their behavior. Accuracy of social perception is largely irrelevant in such situations because it can have little effect upon individuals' behaviors. Indeed, if it does affect individuals' behaviors, it is likely to interfere with role enactment, and, hence, to disrupt the behavior synthesis which is provided by the role system.

REFERENCES

1. ASCH, S. E. *Social psychology.* New York: Prentice-Hall, 1952.
2. BELL, G. B., & HALL, H. E. The relationship between leadership and empathy. *J. abnorm. soc. Psychol.,* 1954, **49,** 156–157.
3. CHOWDHRY, KAMLA, & NEWCOMB, T. M. The relative abilities of leaders and non-leaders to estimate opinions of their own groups. *J. abnorm. soc. Psychol.,* 1952, **47,** 51–57.
4. COTTRELL, L. S., & DYMOND, ROSALIND. The empathic responses: a neglected

---

[2] The writer is at present conducting a laboratory study designed to test this contention.

field for research. *Psychiatry,* 1949, 12, 355–359.
5. DYMOND, ROSALIND. The relation of accuracy of perception of the spouse and marital happiness. *Amer. Psychologist,* 1953, 8, 344. (Abstract)
6. FESTINGER, L., BACK, K., SCHACHTER, S., KELLEY, H. H., & THIBAUT, J. *Theory and experiment in social communication.* Ann Arbor: Institute for Social Research, Univer. of Michigan, 1950.
7. FIEDLER, F. E. Assumed similarity measures as predictors of team effectiveness. *J. abnorm. soc. Psychol.,* 1954, 49, 381–388.
8. GAGE, N. L. Judging interests from expressive behavior. *Psychol. Monogr.,* 1952, 66, No. 18 (Whole No. 350).
9. GAGE, N. L., & SUCI, G. P. Social perception and teacher-pupil relationships. *J. educ. Psychol.,* 1951, 42, 144–152.
10. GREER, F. L., GALANTER, E. H., & NORDLIE, P. G. Interpersonal knowledge and individual and group effectiveness. *J. abnorm. soc. Psychol.,* 1954, 49, 411–414.
11. HEBB, D. O., & THOMPSON, W. R. The social significance of animal studies. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge, Mass.: Addison-Wesley Publishing Co., 1954. Pp. 532–558.
12. HEMPHILL, J. K. Relations between the size of the group and the behavior of "superior" leaders. *J. soc. Psychol.,* 1950, 32, 11–22.
13. HITES, R. W., & CAMPBELL, D. T. A test of the ability of fraternity leaders to estimate group opinion. *J. soc. Psychol.,* 1950, 32, 95–100.
14. MEAD, G. H. *Mind, self and society.* (C. W. Morris, Ed.) Chicago: Univer. of Chicago Press, 1934.
15. MERTON, R. K. Bureaucratic structure and personality. *Soc. Forces,* 1940, 18, 560–568.
16. NEWCOMB, T. M. An approach to the study of communicative acts. *Psychol. Rev.,* 1953, 60, 393–404.
17. NORMAN, R. D. The interrelationships among acceptance-rejection, self-other identity, insight into self, and realistic perception of others. *J. soc. Psychol.,* 1953, 37, 205–235.
18. Ohio State University, Personnel Research Board. *Studies in naval leadership.* Columbus: Ohio State Univer. Research Foundation, 1949.
19. SEARS, R. R. A theoretical framework for personality and social behavior. *Amer. Psychologist,* 1951, 6, 476–484.
20. SPRUNGER, J. A. Relationship of a test of ability to estimate group opinion to other variables. Unpublished master's thesis, Ohio State Univer., 1949.
21. WOOD, H. An analysis of social sensitivity. Unpublished doctor's dissertation, Yale Univer., 1948.

(Received August 3, 1954)

Psychological Review
Vol. 62, No. 4, 1955

THINKING: FROM A BEHAVIORISTIC POINT OF VIEW[1]

IRVING MALTZMAN

University of California at Los Angeles

Hull (10) has demonstrated that the habit family hierarchy and related principles may generate many hypotheses concerning behavior in relation to objects in space such as might occur in the *Umweg* problem or simple kinds of novel behavior. He has thus shown that the elementary laws of behavior may be applicable to behavior of nonspeaking organisms in so-called reasoning situations.

The purpose of this paper is to demonstrate that the principles formulated by Hull and by Spence may also be applicable to the problem solving of articulate humans. In this respect the present analysis has much in common with the important formulations concerning mediated generalization and problem solving by Cofer and his associates (2, 3), Dollard and Miller (5), Doob (6), and Osgood (19). The behavior theory involved may be outlined as follows. Behavior is a function of effective reaction potential ($_{S}\bar{E}_{R}$), which in turn is a multiplicative function of habit strength ($_{S}H_{R}$) and the effective drive state (D) minus the total inhibitory potential ($\dot{I}_{R}$). The latter represents the summation of reactive (I_r) and conditioned ($_{S}I_{R}$) inhibition. It is assumed here that the effective drive state represents the summation of the anticipatory goal response (rg-sg) as well as the primary and secondary drives (23). Furthermore, the multiplicative effect of the anticipatory goal response is restricted to its associated class of instrumental responses.

The principal theoretical conception necessary for our account of problem solving is an extension of Hull's spatial habit family hierarchy (9, 10). The great complexity of human thinking requires the formulation of what might be called compound temporal habit family hierarchies. In the spatial habit family hierarchy, alternative locomotor responses are elicited as a function, in part, of spatial and temporal distance from a goal. But in adult human problem solving, responses in changing spatial relations to a goal are not usually elicited, although there are problems involving motor skills in which this may be the case. A typical performance change in problem solving is in terms of verbal responses, and the change is solely a temporal one (4). Nevertheless, it is assumed that the principles operating in the spatial habit family hierarchy will to a large extent operate in the temporal hierarchy. Recent evidence in support of this assumption has been obtained by Rigby (21).

The conception of a compound temporal habit family hierarchy is based upon the prior assumption that the elementary laws of behavior derived from conditioning and applicable to trial-and-error and discrimination learning are also applicable, at least in part, to primary problem solving or reasoning, and thinking in general. That different kinds of behavior are observed in conditioning, trial-and-error, discrimination, and problem-solving situations is not to be denied. But these different behaviors need not necessarily involve fundamentally different laws. Different behavior is observed in these situations because the initial conditions are different, and the situations represent varying degrees of complexity in the sense of the number of different variables and principles operating in them. Nevertheless, it is reasonable to assume that at least some of the elementary laws derived from conditioning will lead to the development of the composition laws operating in human problem solving.

---

[1] Portions of this paper were presented in a symposium on "Future Trends in Problem Solving" held at the 1954 meetings of the Western Psychological Association.

As Hull (10) has demonstrated, these elementary laws can account for many of the phenomena of simple trial-and-error learning. A hierarchy of responses elicitable by a given stimulus, in which the correct response is relatively low in the hierarchy, characterizes this form of behavior, as shown in Fig. 1. Learning is said to be complete when the order of the response hierarchy has so changed that the correct response is now dominant in the hierarchy. Hull has called this hierarchy of responses, elicitable by a given class of stimuli, the divergent mechanism (9).

As Spence (22) and Hull (10) have demonstrated, the elementary laws of behavior derivable from conditioning situations can also account for many of the phenomena of simple discrimination learning. A hierarchy of stimuli eliciting a given response in which the correct cue is relatively low in the hierarchy characterizes this form of behavior, as shown in Fig. 2. Such learning is said to be complete when the order of the stimulus hierarchy has so changed that the correct cue is dominant. Hull has called this hierarchy of stimuli eliciting a given response the convergent mechanism (9).

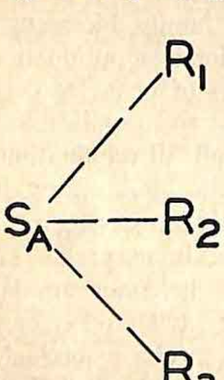


FIG. 1. A divergent mechanism. The stimulus has varying tendencies to elicit the alternative responses.

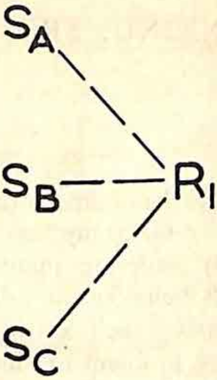


FIG. 2. A convergent mechanism. The alternative stimuli have varying tendencies to elicit a given response.

A synthesis of the divergent and convergent mechanisms gives rise to the habit family hierarchy involved in behavior sequences in relation to objects in space (9, 10). A hierarchy of this sort is shown in Fig. 3.[2]

As seen in this figure, S_A (the external stimulus) and S_D (an internal drive stimulus) are capable of eliciting a given habit family hierarchy and equivalent responses leading to a given

---

[2] There have been a few minor deviations from Hull in the manner of diagramming the stimulus-response relationships in order to simplify their presentation. Instead of having a dashed line between the stimulus and each response member of the divergent mechanism, a single dashed line leading to a bracket is used. All bracketed responses are associated with the stimulus. The number and length of the response sequences have also been reduced. A dashed line between a stimulus and a response signifies a learned association, while a solid line between a response and the cue it produces indicates an unlearned association.

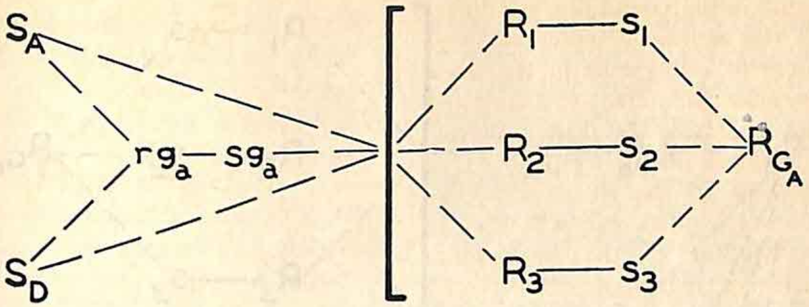


FIG. 3. A habit family hierarchy produced by a divergent and a convergent mechanism.

goal. Common to all responses in the hierarchy is a fractional anticipatory goal response (rg_a-sg_a) which is associated with both the external and internal sources of stimulation. Responses in the hierarchy may be elicited directly by either the external or internal stimuli, or by both. The effects of reinforcement or extinction of individual members of the hierarchy generalize to other members through the mediating mechanism of the anticipatory goal response and its stimulus. It therefore follows that the principles of conditioning and trial-and-error learning should also apply, at least in a general way, to the habit family hierarchy.

For example, if the reaction potential of the correct response leading to a goal is low in the hierarchy, then the generalized conditioned inhibition from the repeated failures of the dominant incorrect responses may reduce its effective reaction potential below the response threshold. Attainment of the goal would never occur under the conditions present. Or, if the subject has a high degree of an irrelevant need such as anxiety, aside from the possible interfering effects coming from the competing responses aroused by this drive, failure to attain the goal under the above conditions would be even more pronounced. The increased effective drive state multiplying the habit strengths for the dominant incorrect responses and the weak correct responses would increase the absolute difference in reaction potential between the two.[3] Such a condition would produce a greater amount of conditioned inhibition generated through the repeated extinction of incorrect responses. This in turn would increase the probability of failure.

Another way in which changes in the habit family may occur is through the arousal of the fractional anticipatory goal response (rg_a-sg_a). Its arousal may produce an immediate increase in effective reaction potential for the related responses. This effect occurs because it presumably enters into a multiplicative relationship with habit strength in the determination of reaction potential.

A synthesis of habit family hierarchies gives rise to the compound habit family hierarchy involved in human problem solving. A hierarchy of this sort is shown in Fig. 4. It is formed when the stimulus of a divergent mechanism becomes a member of

---

[3] For illustrative purposes we may substitute numerical values in the formula for reaction potential ($_SE_R = {_SH} \times D$). The $_SH_R$ value for the dominant incorrect response is 5; the $_SH_R$ value for the weaker correct response is 2; drive has a value of 1. The absolute difference in reaction potential between the correct and incorrect responses is therefore 3. If the drive state is increased to a value of 2, then the absolute difference between responses becomes 6. A greater difference in reaction potential must now be overcome before the correct response can become dominant in the response hierarchy.

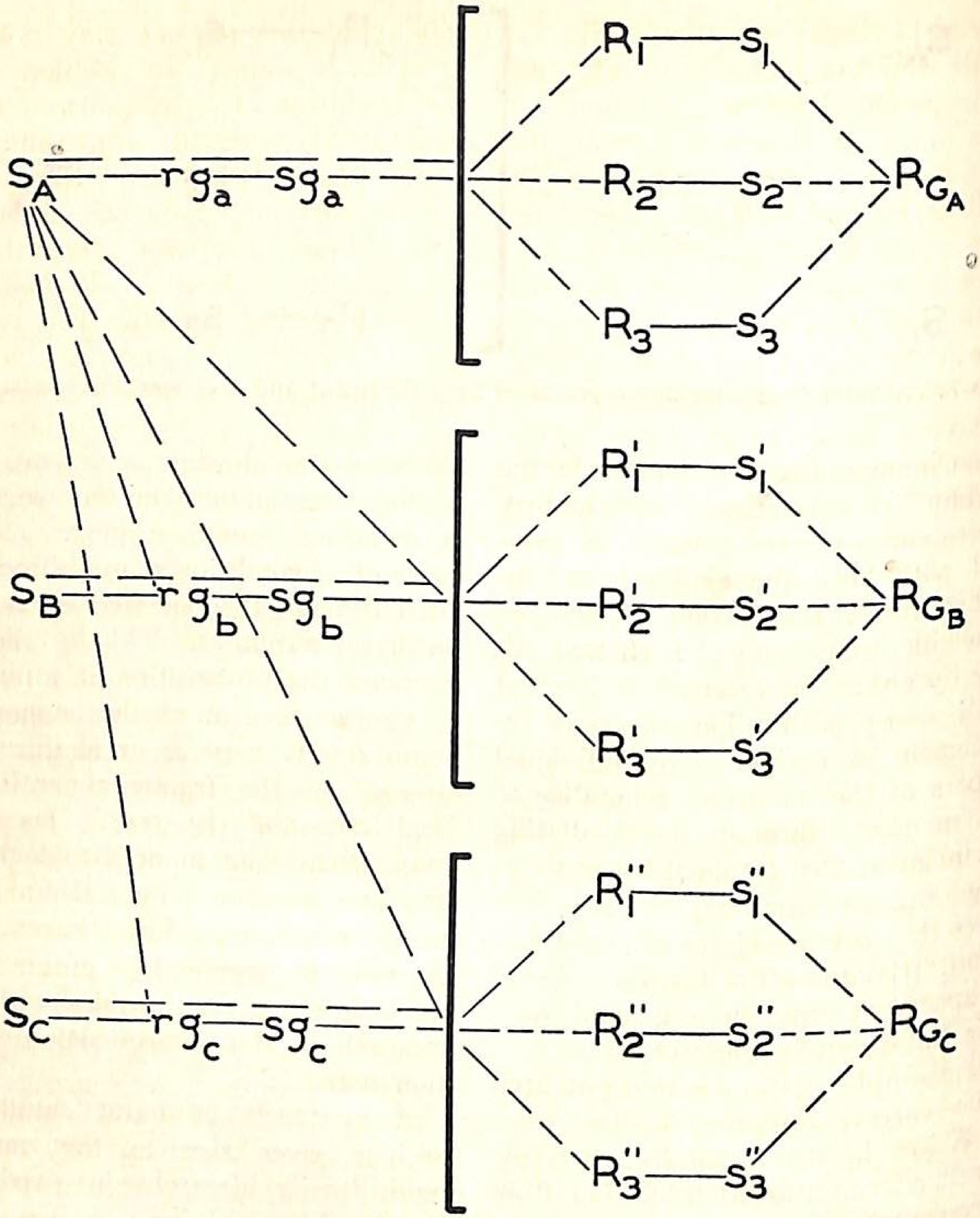


FIG. 4. A compound habit family hierarchy produced by a combination of habit family hierarchies.

a convergent mechanism as well. By the same learning process, responses of the divergent mechanism in question become responses in convergent mechanisms. In the compound habit family hierarchy not only does S_A have the disposition for arousing its habit family hierarchy, but to varying degree the habit family hierarchies of S_B and S_C as well. There is a hierarchy of habit families elicitable by S_A. An analogous condition holds for the other stimulus complexes, S_B and S_C. They have varying amounts of reaction potential for the elicitation of the other habit family hierarchies. These relations as well as the drive stimuli have been omitted in Fig. 4, in order to avoid confusing details.

When the compound hierarchy is formed, R_1, for example, originally only a member of the divergent mechanism elicitable by S_A, becomes a member of a convergent mechanism. There is now a hierarchy of stimuli, S_A, S_B, S_C with differing amounts of effective reaction potential for its elicitation. A similar state of affairs exists for the other response members of a hierarchy, as well as for the anticipatory goal responses. The stimulus complex S_A has varying amounts of reaction potential for the elicitation of the anticipatory goal response rg_b-sg_b and rg_c-sg_c as well as its original anticipatory response. We now have a hierarchy of habit family hierarchies or a class of classes of stimulus-response relationships.[4] Thinking in general, and problem solving in particular, thus may involve the selection of habit family hierarchies as well as the selection of specific response sequences within a hierarchy.

If the selection of response classes or habit family hierarchies obeys the same principles as the selection of individual responses, then the task of discovering the principles of problem solving may be greatly facilitated. The laws derived from conditioning could then be used to account for changes in the compound hierarchy, without, of course, necessarily excluding other principles. Maltzman and his associates have obtained some experimental evidence in support of this assumption (16, 17, 18).

An additional basis for such an assumption is that the difference between instrumental conditioning and problem solving in this respect may not be as great as it seems. In problem solving the members of a response class are qualitatively different—for example, different verbal responses. Effects upon one member may influence other members through mediated generalization, as previously indicated. In instrumental conditioning, changes in a response class, however, are also involved, because precisely the same response may not occur on successive trials (10). There are differences in intensity and, perhaps, in quality in the successive bar-pressing responses in the Skinner box. Nevertheless the rate of bar pressing increases with successive reinforcements. Learning occurs even though a given response is not precisely repeated, because a class of similar responses is reinforced as a result of simple stimulus and simple response generalization. The limits and the precise manner in which generalization occurs in the two situations are probably different. But because of its important theoretical implications for a theory of problem solving, the similarity between the two situations in this respect should not be discounted.

Changes in the order of dominance in the compound hierarchy may occur as a result of either or both of two effects. First, the effective reaction potential of the incorrect dominant habit families and their individual members may be decreased as the result of extinction. The initial response elicited in a given problem situation would tend to be the dominant response in a dominant habit hierarchy. If this response does not lead to a solution, as by definition it would not, it would receive an increment in inhibitory potential, and responses next in the order of dominance would tend to occur. There would be temporary extinction and spontaneous recovery of these incorrect responses. Eventually the inhibitory potential of these responses would reduce their effective reaction potential below that of the responses contained in the hierarchy next in the order of dominance, and so on. Each response in the hierarchy need not be elicited and extinguished, however, since mediated conditioned inhibition may generalize from one member of a hierarchy to another. The anticipatory goal response of this hierarchy presumably would also acquire inhibitory potential, thereby further reducing the effective reaction potential of the related class of responses.

---

[4] This presentation of the compound habit family hierarchy is an oversimplification. For one thing, it does not indicate that individual response members of a habit family may potentially serve as the anticipatory goal response for other habit families. Also, the stimulus aspects of the compound hierarchy, and how they are related to concept formation and perception, are not developed here. Theories of problem solving by Duncker (7), Duncker and Krechevsky (8) and by Wolters (25), although stemming from different points of view, have a number of characteristics in common with the compound habit family hierarchy.

A second general way in which changes in the order of dominance in a compound hierarchy may occur is by increasing the effective reaction potential of the habit family or families initially low in the hierarchy which contain the correct responses. One way in which this may come about is as a result of previous reinforcement of individual members of the hierarchy in that situation. Through mediated generalization all members of the hierarchy would receive an increment in reaction potential.

Another way in which a habit family may be raised in the compound hierarchy is through the arousal of the anticipatory response of the habit family. Elicitation of the anticipatory goal response produces an immediate increase in effective reaction potential for the related class of responses, for the reason previously mentioned (15). The antecedent condition for the anticipatory goal response is assumed to be commerce with a goal or a substitute, often symbolic, for the goal. In adult human problem solving the latter is the typical condition. Recent research indicates that verbal instructions given by the experimenter provide an important condition determining the arousal of the anticipatory response (2, 15). The consequent condition of its arousal is an increased probability of occurrence of responses instrumental in attaining the goal in question.

These different ways in which changes in the compound habit family hierarchy may occur are related to the different kinds of problem solving distinguished by certain writers. As Hull (10) has pointed out, any learning situation other than classical conditioning involves problem solving of some sort. The essential characteristics of a problem-solving situation are that an organism is motivated, and that attainment of some goal object satisfying that drive is dependent upon the organism's performing in a given manner. It is characteristic of problem solving that the appropriate response leading to goal attainment does not immediately occur. This is true to some extent in instrumental conditioning, and to a greater extent in trial-and-error learning and in what has traditionally been called problem solving.

Further distinctions between different kinds of problem solving have also been made. Maier (13, 14) has repeatedly distinguished between different functions responsible for problem solving, and has classified problem solving as either productive or reproductive. The latter kind of thinking according to Maier involves the application of previously acquired experiences which lead to a correct solution in a new situation. It is problem solving based on the transfer of training, or equivalent stimuli. Productive thinking on the other hand is the consequence of the integration of previously unrelated experiences. The integration is produced by a direction which is an "outside" force not itself a habit.

The present treatment of problem solving, as previously indicated, makes use of the concept of the fractional anticipatory goal response. Certain functional characteristics of the anticipatory goal response appear to make it an analogue of Maier's concept of a new direction. However, among other things, we do not accept Maier's restriction that only productive thinking involves the combination of previously isolated habit segments, to use Hull's terminology. To some extent this is the distinguishing characteristic of all thinking. It is the feature that sets it off from simple retention, and trial-and-error learning. All forms of thinking involve mediated generalization, and hence compounding of previously isolated habit segments.

Nevertheless, there are differences between reproductive and productive thinking. Situations eliciting reproductive thinking often involve the presentation of a succession of problems. The solution of each of these requires the elicitation of different response members of the same habit family hierarchy. This habit family will then become dominant in the compound hierarchy, as the result of reinforcement of its response members and the extinction of responses belonging to different habit families. As previously indicated, the increase in reaction potential of the entire class of responses as the result of reinforcement of individual members would occur as a consequence of mediated generalization. For example, subjects given a series of anagrams whose solutions all belong to the same word category will have a greater frequency of success on subsequent anagrams of the same category than subjects without such prior experience (20). However, failure on individual problems may occur because a solution is still dependent upon selection of particular responses within this one habit family.

The occurrence of reproductive thinking in situations where only a single problem is presented is also the consequence of the factors outlined. As a result of past training the dominant habit family in the compound hierarchy contains the correct response. Solution of the given problem depends upon extinction of the initially dominant incorrect responses within this one hierarchy.

In productive thinking, on the other hand, a habit family initially low in the compound hierarchy must become dominant before a correct solution can be attained. This occurs as the result of the extinction of the dominant incorrect response hierarchies. Once the appropriate habit family is dominant a solution will occur, provided that the correct responses within that hierarchy in turn become dominant. The protocols of subjects in Duncker's radiation problem exemplify this mode of thinking (7).

If a subject incorrectly anticipates the goal or solution to the problem—e.g., "destroy the tumor by means of rays sent over a path as free as possible from healthy tissue"—the proposed solutions are alternative responses within the same dominant habit family, and are all attempts to achieve this end. Repeated failure of these proposals will produce extinction of the anticipatory goal response and its related instrumental responses. As a consequence, another habit family may become dominant. The subject will now anticipate a different kind of solution, such as "reduce the intensity of radiation." As a result, all potential response sequences in the subject's habit family leading to this goal are facilitated. If the correct solution (converging rays from different angles) is not the dominant response in this hierarchy, extinction of incorrect responses must occur before the solution will be attained.

In an actual protocol this orderly progression from one habit family to another is probably an infrequent occurrence. The more typical case would be one in which the reaction potentials of two or more habit families overlap. This would probably be the case after the first few responses from the dominant hierarchy have been extinguished. Thus reproductive and productive thinking differ with respect to the kinds of changes which must occur in the compound hierarchy before a correct solution can be attained. In reproductive thinking the habit families containing the correct responses are dominant at the outset of the problem, as the result of training and generalization from other situations. Or they rapidly become dominant in the compound hierarchy through the reinforcement of individual response members of the habit family. In productive thinking the habit families containing the correct responses are initially low in the compound hierarchy. They become dominant following the extinction of the dominant habit families which lead to incorrect solutions.

Experiments on the effects of direction in the pendulum problem by Maier (12) and by Weaver and Madden (24) are a special case of productive thinking, in that instructions and demonstrations are employed to increase the reaction potential of the anticipatory goal response and individual response members of the habit families leading to a correct solution. The problem is to construct two pendulums which would make chalk marks on two different places on the floor. In Maier's experiment (12) one group of subjects was given only the statement of the problem. Other groups received the statement of the problem plus various additional instructions or demonstrations. One of these groups was given three different demonstrations of operations on the material which were necessary for solution of the problem. They were shown how to make a plumb line, how to combine poles by using a clamp, and how to wedge poles against a surface. A third group was given these demonstrations and told that they must combine them for a solution of the problem. A fourth group was told that it would be advantageous if the pendulum could hang from the ceiling. They were given a "direction." The fifth group received the demonstrations and the direction. All of the problem solutions except one occurred in this last group.

According to the present formulation the experimenter's statement that it would be advantageous to hang the pendulum from the ceiling tended to elicit an anticipatory response for this goal. A wide variety of equivalent responses instrumental in leading to this goal therefore received an immediate increment in reaction potential. However, for a correct solution to occur, certain specific responses must be elicited. The three demonstrations given the last group increased the tendency of these responses to occur within their respective habit family hierarchies. The increased frequency of solutions under these conditions would follow from the differential increase in reaction potential of the relevant responses and the lawful nature of trial and error learning. However, since a large number of responses belong to a given hierarchy, and the correct responses may still not be dominant, extensive extinction of incorrect responses must occur. If the correct responses are very low in the dominant hierarchy, a solution may not occur at all because of the extensive generalization of the effects of extinction of the incorrect responses. Another basis for failure in this group is that despite the instructions tending to arouse the appropriate anticipatory goal response, some subjects, presumably as the result of self-instructions, induce different anticipatory goal responses. These are the subjects that adopt an inappropriate approach to the problem, according to Maier.

Groups 2 and 3, which are not given the directional instructions but receive the demonstrations, also have an increase in reaction potential for the three response sequences necessary for a solution. However, the anticipatory goal response and the related class of responses necessary to suspend the pendulum from the ceiling are not increased in strength, which is presumably why these subjects failed to solve the problem. However, contrary to Maier, Weaver and Madden (24) found no difference between the performance of groups 3 and 5.

Their experiment implies that the appropriate habit family for suspending the pendulums from the ceiling may be elicited by self-instructions, or that the increase in habit strength of the three necessary response sequences by itself may be sufficient for the solution of the problem in some individuals. Why there was this discrepancy between the two studies, however, is not at all apparent.

Throughout the previous discussion the systematic and theoretical status of the concept of thinking has only been implied. We shall now try to make it more explicit. According to the present systematic position, thinking is a defined concept or hypothetical variable. The specific definition that is given to it is the problem of theory, and will be discussed presently. Now we must explore further the consequences of the assumed systematic status of the concept. For one thing, thinking is not a response, verbal or otherwise, just as learning is not a response, and just as electricity is not the temperature of a conductor. All of these are dispositional concepts that are given empirical meaning by statements referring to their antecedent as well as their consequent conditions (1). They are not equivalent to their manifestations or consequent conditions. The insistence that thinking *is* a verbal response, a contraction of certain muscles, or activity in the central nervous system, is thus based on an inappropriate use of language. The verbal response, for example, is just one of several different kinds of responses that may be taken as a criterion or manifestation of thinking. Other response criteria might be gestures, mimicry, motor skills, etc. Questions as to which ones may be taken as criteria, and under which conditions, as well as how they are related to thinking, are to be answered by experiments and theory.

The present systematic position with respect to the relationship between thinking and verbal responses (the most frequently used criterion of thinking) may be made clearer by using the analogy of bar pressing in the Skinner box. The assumptions here are analogous to the assumption that bar pressing is a function of other variables besides learning. If a bar depression does not occur, it need not necessarily imply the absence of learning. The rat's motivation may have been reduced to a minimum; there may be temporary extinction of the response, or inhibition due to the arousal of competing responses by extraneous stimuli, etc. Similarly, the absence of a verbal or some other kind of response does not necessarily mean the absence of thinking. It may be due to the absence of effective reaction potential for that particular response; perhaps the relevant motivation is absent; or other response tendencies inhibit its appearance, etc. On the other hand, the presence of bar pressing does not necessarily imply that learning has occurred. It may be an operant level, some unconditioned response strength—in Hull's terminology, ${}_{S}U_{R}$. Likewise, verbal responses may occur in the absence of thinking.

One aspect of this condition will be discussed shortly.

In the foregoing we have tried to explicate the systematic status of the concept of thinking. We shall now turn to the more specific problem of how it may be treated within the framework of Hull's theory of behavior.

Since it is assumed that thinking as well as learning is a disposition or hypothetical variable, the problem now is to distinguish between the two concepts. If the two are defined in terms of the same operations and consequences, they have the same empirical meaning and the distinction is purely a verbal one. As commonly employed, the term learning refers to the acquisition of a hypothetical state, ${}_{S}H_{R}$, as a result of antecedent conditions such as the number of reinforcements. The consequent conditions are changes in some response criteria such as decreased latency, increased rate of responding, etc.

The term thinking as it is employed here refers to the utilization of new combinations of habit strength by articulate organisms. In other words, we assume that thinking is equivalent to a complex form of effective habit strength which is produced by mediated generalization. The reason for arbitrarily restricting the usage of "thinking" to humans is the belief that extensive mediated generalization is necessary for the recombination of habit strengths to occur, and complex mediated generalization of this sort is made possible primarily by linguistic responses.

We have stated that thinking involves the utilization of learning in new combinations, as distinguished from the acquisition of learning. A further problem is to distinguish between thinking and retention, since the latter also involves the utilization of habit strength.

The distinction is not always easy to make, and at times may be arbitrary. But so is the commonly accepted distinction between learning and retention (11). As previously noted, learning refers to the acquisition of a hypothetical state as a function of the number of reinforcements. Retention is a term referring to the persistence and subsequent manifestation of that hypothetical state. In a learning experiment, performance on trials after the first is a function of the persistence of previous learning, or retention, from earlier trials. Only on the first trial is nothing more than the acquisition of habit strength involved. This entire process, however, is called learning even though much of it is actually retention. After the subject has reached some predetermined criterion, he is required to utilize in some manner the habit strengths previously acquired. He is asked to recall the material previously learned, recognize, or relearn it, etc. In every case the responses originally acquired are elicited again to some extent. This implies that the habit strengths utilized in the test of retention are substantially the same as that originally acquired. Retention has as its consequent condition the elicitation of some previously acquired response, presumably as a result of the persistence of previously acquired habit strength. The term thinking has as its consequent condition the elicitation of a response other than the previously acquired response as a result of past learning. Habit strength previously acquired has entered into new compounds, has changed as the result of mediated generalization.

A fundamental problem for a behavioristic psychology of thinking is to determine the laws governing these combinations and recombinations of habit strengths. Hull's equations (10) for combining habit strengths in generalization and compound stimulus situations are approximations of such composition laws. But they are only first approximations, because the generalization and compounding of habit strengths occurring in thinking are undoubtedly a good deal more complex than those that Hull has treated. A basic problem in this respect would be the development of the laws of mediated generalization which, theoretically, produce the new compounds of habit strength, and empirically, produce the formation of new stimulus-response classes. It is likely that the close connection between language and thinking (or even their equating) in certain theories results from the fact that language permits the greatest degree of mediated generalization and therefore thinking.

Admittedly the theory of thinking and problem solving outlined here is loosely formulated and incomplete. Nevertheless, it at least has the merit that it relates human problem solving to behavior in simpler situations. It is an attempt to integrate the two, as distinguished from the usual gestalt approach which treats problem solving as divorced from the relatively large number of principles derived from conditioning and trial-and-error learning. Although many of the principles derived from conditioning may not entirely apply to human problem solving, this is certainly an empirical question worth investigating. At the very least, these principles should yield significant hypotheses as to the factors determining problem solving.

SUMMARY

A theory of human problem solving has been outlined, based upon the concept of a compound temporal habit family hierarchy, which is assumed to function, at least in part, according to the principles of conditioning and trial-and-error learning. Some of the characteristics of the compound hierarchy were noted, and its role in different kinds of problem-solving situations was indicated.

The systematic status of thinking from a behavioristic point of view was described as a disposition or hypothetical state of the organism. Within the present theory it is equivalent to a new combination of habit strengths produced, primarily, by mediated generalization.

REFERENCES

1. CARNAP, R. Logical foundations of the unity of science. *Int. Encyc. unif. Sci.*, 1938, **1**, 42–62.
2. COFER, C. N. Verbal behavior in relation to reasoning and values. In H. Guetzkow (Ed.), *Groups, leadership and men*. Pittsburgh: Carnegie Press, 1951. Pp. 206–217.
3. COFER, C. N., & FOLEY, J. P., JR. Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.*, 1942, **49**, 513–540.
4. COHEN, J. The concept of goal gradients: a review of its present status. *J. gen. Psychol.*, 1953, **49**, 303–308.
5. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw Hill, 1950.
6. DOOB, L. W. The behavior of attitudes. *Psychol. Rev.*, 1947, **54**, 135–156.
7. DUNCKER, K. On problem solving. *Psychol. Monogr.*, 1945, **58**, No. 5 (Whole No. 270).
8. DUNCKER, K., & KRECHEVSKY, I. On solution-achievement. *Psychol., Rev.*, 1939, **46**, 176–185.
9. HULL, C. L. The concept of the habit-family hierarchy and maze learning. *Psychol. Rev.*, 1934, **41**, 33–54, 134–152.
10. HULL, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.
11. MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning*. New York: Longmans, Green, 1952.
12. MAIER, N. R. F. Reasoning in humans: I. On direction. *J. comp. Psychol.*, 1930, **10**, 115–143.
13. MAIER, N. R. F. The behavior mechanisms concerned with problem solving. *Psychol. Rev.*, 1940, **47**, 43–58.
14. MAIER, N. R. F. Reasoning in humans: III. The mechanism of equivalent stimuli and of reasoning. *J. exp. Psychol.*, 1945, **35**, 349–360.

15. MALTZMAN, I., & EISMAN, E. Two kinds of set in problem solving. Paper read at Amer. Psychol. Ass., New York, September, 1954.

16. MALTZMAN, I., FOX, J., & MORRISETT, L., JR. Some effects of manifest anxiety on mental set. *J. exp. Psychol.,* 1953, **46**, 50–54.

17. MALTZMAN, I., & MORRISETT, L., JR. Different strengths of set in the solution of anagrams. *J. exp. Psychol.,* 1952, **44**, 242–246.

18. MALTZMAN, I., & MORRISETT, L., JR. The effects of single and compound classes of anagrams on set solutions. *J. exp. Psychol.,* 1953, **45**, 345–350.

19. OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

20. REES, H., & ISREAL, H. An investigation of the establishment and operation of mental sets. *Psychol. Monogr.,* 1935, **46**, No. 6 (Whole No. 210).

21. RIGBY, W. K. Approach and avoidance gradients and conflict behavior in a predominantly temporal situation. *J. comp. physiol. Psychol.,* 1954, **47**, 83–89.

22. SPENCE, K. W. The nature of discrimination learning in animals. *Psychol. Rev.,* 1936, **43**, 427–449.

23. SPENCE, K. W. Theoretical interpretations of learning. In C. P. Stone (Ed.), *Comparative psychology.* New York: Prentice-Hall, 1951. Pp. 239–291.

24. WEAVER, H. E., & MADDEN, E. H. "Direction" in problem solving. *J. Psychol.,* 1949, **27**, 331–345.

25. WOLTERS, A. W. On conceptual thinking. *Brit. J. Psychol.,* 1933, **24**, 133–143.

(Received July 8, 1954)

Psychological Review
Vol. 62, No. 4, 1955

VISUAL FIGURE DISCRIMINATION AND THE MEDIATION OF EQUIVALENCE RESPONSES

A. L. TOWE

Department of Physiology and Biophysics, University of Washington School of Medicine

Among the most prominent features of any group of discrimination performance data are (*a*) the extensive daily variations in performance level of any subject and (*b*) the differences among subjects in the total number of training trials required to attain a specified criterion performance level. Psychologists have devised several methods designed to minimize these characteristics and thereby to display a continuous function designated a "learning curve." The derived function is very often thought to be a "curve of learning." Such a manipulation seems implicitly to ignore, or even to deny, the tenet in psychology that all behavior is determined, and to recognize certain behaviors as "chance or random" much after the manner of Aristotle. The psychologist, as a behavioral scientist, ought not to slur over this inconsistency in behavior, but rather should deal with the problem in his thinking and his theorizing. One attempt to face up to the problem is tendered here without further apology.

In 1932, Krechevsky (2, 3), extending Lashley's notion of "attempted solutions" (4), endeavored to account for discrimination learning in terms of "hypotheses" during the presolution period. The basic supposition of the notion is that, so long as no hypothesis involves the differential cue stimuli, associations are not formed to these cue stimuli, and hence the discrimination problem is not solved. Errors following solution of the problem are irrelevant to the formation of the association, and are not adequately accounted for in this system. Spence, on the other hand, believed that discrimination learning could be handled in terms of trial and error learning theory, and in 1936 (7) published a theoretical formulation based upon the principles of reinforcement and nonreinforcement of discriminably different stimuli, making learning a continuous process. Hypothetical generalization curves were presented (9) which showed performance at any time during training as a function of the interplay of excitatory and inhibitory tendencies with respect to the cue stimuli. It was made explicit that weakening, or inhibition, follows nonreinforcement only when the stimulus already has a certain minimum excitatory strength (8). The notion does not adequately explain the magnitude of the variability seen in discrimination performance data, for it predicts relatively uniform increments in performance level during successive training periods.

Two Assumptions

Because of the writer's bias toward a continuity theory, an attempt will be made to extend this type of formulation to account for both the wide variability in performance from one training session to the next and the differences in rate of acquisition, and also to account for the character and extent of transfer of an acquired visual figure discrimination to variations from the training figures. The idea was developed during a discrimination study with pigeons (10), but tentatively will be treated as though descriptive of mammalian behavior as well. Two assumptions are necessary to the formulation, one seeming axiomatic and the other being less obvious and having no direct experimental support. The first assumption is that any visual figure may be perceived in several different ways, i.e., a visual figure has several aspects to which an organism may learn to react. Secondly, it is assumed that the organism does not perceive the same visual figure in the same way on all trials or training sessions, i.e., the organism does not always react to the same aspect of the stimulus from one trial or training session to the next.

The term "aspect," as used above, requires some discussion. Aspect will *connote* a unitary and organized perception involving a portion or property of a visual figure as "figure." Aspect will *denote* that portion or property of a visual figure which is necessary and sufficient for a differential conditioned response by a particular organism, when confronted with that portion or property alone or in conjunction with one or more additional visual figures or portions or properties of them. Thus if, after a positive response to a triangle and avoidance of a square were established, a line of the same length and orientation as the base of the triangle evoked a positive response when appearing alone or with the square, or any other figure, and would not evoke a positive response when its length or orientation was altered, then that line in that orientation would be an aspect. Similarly, any other part of the triangle, e.g., an angularity property, which evoked a positive response when presented alone would be an aspect. Aspects are additive, divisible, or reducible for any organism; they need not be absolute properties, but may also be relative properties among figures. It must be pointed out that these aspects, in contrast with Hullian elements, are not necessarily the units which a physicist may find, but are momentarily unitary with respect to the perceiving and reacting organism. In short, there are certain absolute properties in any visual figure and certain relative properties among visual figures to which an animal could learn to respond, which will then be called aspects or stimulus aspects. Hebb (1, p. 105) has developed a notion very much the same as this in his schema.

If an animal perceives a portion or property of both the positive and the negative figures in the same way, an ambiguous aspect emerges. Those portions or properties of the figures giving rise to the same perceptual response may or may not be physically identical. If they are not identical in the physical sense, the animal may at some future time differentiate them (5, 6). When differentiation occurs, the "new" aspect of the negative figure may now acquire an inhibitory tendency. That is, when an excitatory value is attached to an ambiguous aspect which is subsequently perceived as two different aspects, an inhibitory tendency may be built up to the new aspect of the negative figure (after Spence), and performance will then show improvement of a step-wise nature.

SOME PREDICTIONS

Because the mere germ of the idea has been sketched here, a few predictions deriving from it might elucidate the theory. Several relations follow from the above formulation.

1. As the criterion performance level for initial learning is raised, the amount of transfer to new figures will increase. The variety of visual figures to which an animal responds, as equivalent to the original positive training figure, depends upon the number of positive and negative aspects that the animal has acquired during training. The longer the subject is exposed to the training figures, the greater will be the number of aspects acquired. It follows that

transfer will be more extensive with greater amounts of training.

2. As training progresses, the ratio of adient to avoidance responses will change. In the initial stages of learning, if no transfer from prior experience is involved, all responses should be adient in nature. As training progresses and ambiguous aspects are split into positive and negative aspects, there should be an increasing number of responses directed largely by avoidance of certain aspects. This visual avoidance of negative aspects will increase the probability that the animal will respond to a positive aspect of the other figure; it is believed that the negative stimulus aspect does not itself result in the conditioned behavior of responding to the "other" figure. The work of Lawrence (6) also hints at this sort of mechanism. The portion or property of the figure which has no associated inhibitory value will not direct the visual exploration of the animal away from the "negative" figure, and thus should play no role in transfer behavior.

3. The amount of training required to attain a specified criterion will vary with the complexity of the figures employed. Very simple figures have relatively few potential aspects, i.e., will be perceived in a limited number of ways. When few aspects are involved, performance should improve quite rapidly. If the rate of "shopping about" is independent of figural complexity, then the total number of trials required to attain a certain performance level will increase with increases in the number of ways the animal perceives parts of the figure.

4. The ratio of adient to avoidance responses will vary with the complexity of the figures used. Avoidance responses will be least for simple and grossly different figures, and will increase as the complexity and similarity of the figures increases (increasing the probability that ambiguous aspects will develop). Changes in the ratio of ambiguous to positive and negative aspects have significance for the problem of experimental neurosis.

This kind of formulation can be extended outside the realm of visual figures to account for the form of the generalization gradient and such phenomena, but will not be attempted here. It should be pointed out that different species will probably display various degrees of the tendency to "go shopping"; the extent of transfer to altered figures will be related to this tendency. Whether or not a particular figure is responded to positively depends upon how that figure or a part of it is perceived at the moment.

REFERENCES

1. HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.
2. KRECHEVSKY, I. 'Hypotheses' versus 'chance' in the pre-solution period in sensory discrimination-learning. *Univer. Calif. Publ. Psychol.*, 1932, **6**, 27–44.
3. KRECHEVSKY, I. The genesis of 'hypotheses' in rats. *Univer. Calif. Publ. Psychol.*, 1932, **6**, 45–64.
4. LASHLEY, K. S. *Brain mechanisms and intelligence.* Chicago: Univer. Chicago Press, 1929.
5. LAWRENCE, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39**, 770–784.
6. LAWRENCE, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40**, 175–188.
7. SPENCE, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.
8. SPENCE, K. W. Analysis of formation of visual discrimination habits in the chimpanzee. *J. comp. Psychol.*, 1937, **23**, 77–100.
9. SPENCE, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.
10. TOWE, A. L. A study of figural equivalence in the pigeon. *J. comp. physiol. Psychol.*, 1954, **47**, 283–287.

(Received August 9, 1954)

Psychological Review
Vol. 62, No. 4, 1955

ELICITATION THEORY: I. AN ANALYSIS OF TWO TYPICAL LEARNING SITUATIONS

M. RAY DENNY AND HARVEY M. ADELMAN
Michigan State College

The present paper will be concerned with an acquisition hypothesis and one or two other elicitation principles which constitute part of a general theory of behavior now under development.[1] However, there will be no attempt to present the theory here in a formal or definitive manner. Our purpose is to indicate the operation of the theory by applying it in the analysis of two types of instrumental learning, namely, in a conventional **T** maze and in a Skinner box. Simple approach learning in a runway will not be discussed because its explanation can be interpolated from the analysis of the more complex situations.

THE ACQUISITION HYPOTHESIS

The acquisition hypothesis, which has its roots in both Guthrie's contiguity principle and Hull's reinforcement principle, may be stated as follows: (*a*) The stimulus complex (S) which closely precedes in time any response *elicited* by any stimulus (S_e) acquires the property to elicit this response. (*b*) With each elicitation there results an increment to the tendency of the stimulus complex (S) to elicit this response. (*c*) The S-R association will not ordinarily be evident in behavior unless the response is *consistently* elicited in the given stimulus situation ($S + S_e$). Thus, for all practical purposes, learning occurs only when a response is prepotent over a series of trials or over an extended period of time.

Two terms require explication. By *elicitation* is meant that the response is evoked in some manner which necessarily involves the afferent nervous system. We use *consistently* to refer to whether or not the response is elicited each time the stimulus (S) is present: The more often R is elicited each time S is present, the more consistent the elicitation.

In other words we are postulating that the essential condition for the strengthening of one response tendency over and above another response tendency is the *consistent* elicitation of the response in question. For example, the US of a classical conditioning experiment typifies a consistent elicitor. Here, it is appropriate to point out that conditioning of a specified response will not occur until all irrelevant elicitors have been eliminated. For example, Pavlov's dogs could not be conditioned until they were habituated to the total experimental situation, that is, not until the "neutral" stimuli in the conditioning situation had lost their eliciting property through a process of stimulus satiation or adaptation (7, 9).[2]

---

[1] The present theory in many respects is parallel to a theory under development by Maatsch (14) in the same laboratory. It has similarities to ideas currently being expressed by Nissen (16, 17), and by Sheffield and others (20, 21).

[2] The principle of sensory elicitation and adaptation is fundamental in the present system. This principle refers to the fact that though all stimuli have some initial eliciting property most of them soon lose or partially lose this property through continued or repeated presentation. The stimuli most resistant to adaptation are the common reinforcing agents (food, water, shock, etc.). So-called neutral stimuli constitute the class of stimuli which commonly evoke investigatory, exploratory, and manipulatory responses (7, 9, 15). Such responses fall into the more general cate-

Another example of a consistent elicitor as found in an instrumental learning situation is familiar food for a hungry animal or water for a thirsty animal. In this case the approach to the goal object and the consummatory response are the responses consistently elicited. Since in a maze these responses are usually the only ones consistently elicited, they are the ones which are conditioned to the cues of the goal area.

Noxious stimulation is also a consistent elicitor and produces learning of the responses elicited to the contiguous stimuli, e.g., escape, fear, and avoidance.

Derived or Secondary Elicitation

There remains at least one more important class of eliciting agents. Once a behavior sequence has been established to some degree and the response no longer leads to the goal object or is no longer followed by the *US,* then there exists a condition for the elicitation of a new class of characteristic responses. This is to say, for example, that nonreward or frustration, either terminal or momentary, constitutes an eliciting (reinforcing) state of affairs.[3] In experiments involving nonreward it is customary to observe the characteristic response of recoil after entering a cul or negative end box, and of attempting to escape the place where reward was previously obtained (5). Since the omission of a goal object consistently elicits avoidant-type responses, and since the elicitation of these characteristic responses constitutes an instance of reinforcement,[4] the cues associated with nonreward acquire the property to elicit avoidance. With continued strengthening these avoidant responses compete with the original response and extinction of the approach response results.

We also postulate that the omission of noxious stimuli will elicit responses antagonistic to the original response. In general the responses which arise through the withdrawal of noxious stimuli belong to a class of "relaxational" or approach responses. Consistently elicited relaxational-approach-type responses are therefore conditioned to stimuli associated with the withdrawal of a noxious stimulus, and with continued strengthening compete with the original response and produce extinction. Experimental evidence in support of this notion may be found in two recent studies (4, 22). In a subsequent paper we hope to show that the eliciting property, rather than the "drive reducing" property, of the cessation or omission of shock is the *sine qua non* of typical instrumental-avoidance learning.

In summary, secondary elicitation, i.e., the eliciting of response by the omission of the *US* in a previously established behavior sequence, constitutes an instance of reinforcement just as the elicitation of any other response by an

---

gory of approach behavior. With the passage of time all stimuli recover their capacity to elicit responses, approach or otherwise. One result of this phenomenon is variability in behavior, e.g., spontaneous alternation.

[3] In this connection it should be noted that Amsel and Ward (3) have demonstrated the cue properties of frustration, and Amsel and Roussel (2) have demonstrated the eliciting (drive) properties of frustration. In addition, Goer (8) has shown that as few as 12 trials reinforced on the same side as the rat's original position habit will facilitate the learning of a subsequent black-white discrimination. Presumably, the shift from consistent reward of the position response to 50 per cent reward sets up secondary elicitation and mediates the learning of avoidance reaction to positional cues.

[4] Reinforcement in the present context does *not* refer to the strengthening of a closely *preceding* response in the Hullian or Thorndikian sense. Reinforcement occurs simultaneously with the eliciting of the response, that is, it *is* the eliciting of a response. In similar fashion, a secondary reinforcing stimulus is simply a discriminative stimulus or conditioned elicitor.

US constitutes an instance of reinforcement.[5]

APPLICATION

Since the present position posits that all learning is fundamentally classical conditioning,[6] the task is to explain all instrumental learning within this framework. To this end we will analyze two typical instrumental learning situations.

Learning the **T** *Maze*

Let us assume that we are attempting to train a group of rats to turn (right) into arm A of a conventional **T** maze where the extra-maze stimuli are discriminative. All *S*s have been placed on a reduced food regimen and handled for one week; no pretraining has been given; and a small pellet of food will serve as a reward.

Correction Technique

S is placed in the starting box. The novel stimuli of the **T** maze then elicit exploratory (approach) response, and *S* will eventually enter the positive end box at the end of arm A where approach and eating responses will be elicited and thus conditioned to the cues of this box. Somewhere near this point, stimulus generalization becomes an important factor. Thus the to-be-negative arm (arm B) which is similar to arm A will also acquire the tendency to elicit approach response. This, taken in conjunction with the fact that the less recently visited stimuli of arm B will possess greater exploratory value (approach elicitation), increases the probability of *S*'s entering arm B on some succeeding trial.

When *S* does enter the negative end box (arm B), the *US* (food) is not present. Therefore, secondary elicitation obtains, and the conditioning of avoidant behavior is initiated. Since in the correction technique recoil or escape responses can clearly occur after each wrong response, these responses will become conditioned to the cues of arm B, and eventually these cues will become signals of avoidance rather than approach. With continued trials (reinforcement) the cues of the choice-point region consistently elicit avoidance to B and approach to A. Thus, T-maze learning within this framework is regarded as a case of discrimination (perceptual) learning in which the stimuli of the choice-point region acquire the discriminative property to elicit avoidance responses to one arm and approach responses to the other.

---

[5] This means that the laws of acquisition apply to extinction as well as to original learning. The only difference is with respect to the eliciting agent and how it functions. Thus it is quite possible that certain conditions which favor original learning may not at all favor the learning which takes place during extinction, and vice versa (1). The net result is to conceive of extinction as an interference phenomenon. The response tendency established during extinction (on the basis of secondary elicitation) competes with the response tendency of acquisition and eventually predominates, at least temporarily.

Discrimination is also involved. Any alteration of the stimulus conditions when shifting from acquisition to extinction, as in terms of spaced vs. massed trials, will facilitate the extinction process. If, as is typical, acquisition is spaced and extinction is massed, the introduction of a rest interval will tend to reinstate the "spacing" cues of acquisition and remove the "massing" cues of extinction. The net result is "spontaneous recovery."

In this interference theory of extinction there is no gradual unhooking of responses, as in Guthrie, and there is no weakening of the original response tendency by "non-reinforcement." It is simply that for *similar* stimulus situations there are two or more competing response tendencies.

[6] It should be made clear that in considering all learning to be a matter of classical conditioning we are not drawing a strict parallel between conventional *CR* experiments and simple approach or avoidance *CR*'s established during instrumental learning. The conditioning of an eyelid *CR,* for example, involves considerable discrimination learning in terms of specifying the *CR* to the momentary *CS* and *not* to the stimulus complex in general. Since typical *CR* experiments necessarily involve discrimination (18), *S* must learn to make some response other than the *CR* when the *US* is omitted between trials. If secondary elicitation is minimal or weak, as would be the case during the omission of a puff of air to the cornea, or a blow on the patellar tendon, or a bright light to the pupil, then many typical *CR*'s should be difficult to set up. Furthermore, in many typical *CR* situations, particularly for those using human *S*s with uncontrolled attitudes, the *US* may not be the prepotent stimulus on every trial, and as a result conditioning never reaches a high level (12, 19). However, when a dominant elicitor such as food for a hungry animal is employed, both the presence and omission of the goal object are prepotent elicitors on every trial. Therefore, the conditioning posited to account for instrumental learning is rapid and rises to a high level.

In the context of the present theory it is necessary to point out that the responses (exploratory) elicited by novel stimuli will not be conditioned to the maze cues in a differential manner.[7] These responses are not consistently elicited by one particular side of the maze. Rather, on the basis of exploratory considerations alone, one would expect spontaneous alternation (6, 7, 15). However, the avoidance tendency conditioned to arm B will compete with the exploratory approach tendency and will virtually preclude entrance into that arm after a large number of trials.

Noncorrection Method

The explanation of learning a T maze with the noncorrection method is essentially the same as with the correction method. However, the fact that *S* is not allowed to recoil from the negative end box and also enter the positive end box on the same trial in the noncorrection method makes it necessary to treat these two training methods separately. Presumably, detention in the negative end box decreases the probability that any one response will be consistently elicited or specified to the cues of the negative end box. Also the cues of the choice point area in arm B are not directly associated with escape responses, i.e., there is greater *S-R* asynchrony than with the correction method. It therefore follows that the learning of avoidance to the cues of arm B will be appreciably slower than with the correction method and that learning of the correct response with the noncorrection method accordingly will be slower (10). Avoidance learning to the negative side, although less than with the correction method, nevertheless does occur. It has been found, for example, that the learning of a T maze (noncorrection method) is enhanced when the number of rewards on the correct side is held constant (5).

Place vs. Response Learning

From the point of view of the present theory there is no *real* controversy as to whether *S* learns a right-turn response or learns to go to the correct place (arm A). The concept of learning a turn response can only mean that the cues directing the response remain constant regardless of the orientation of the stem with respect to the arms. And in any T-maze situation a number of tactual, kinesthetic, and visual patterning cues in the choice-point area remain constant and thus mediate response learning when the cues beyond the choice point are nondiscriminative. Once *S* has made a response at the choice point under pure response learning conditions, there are no longer any stimuli in the maze which are discriminative other than a perseverative stimulus trace. For the cues of each arm, including extra-maze stimuli, will have acquired identical approach (avoidance) value. In Hullian terms this means that pure response learning involves delay of reinforcement; in the present context we would prefer to say that pure response learning involves a long *CS-US* interval, i.e., a lapse of time between the cessation of the discriminative stimulus and approach-to-food. As previously implied, any deviation from strict contiguity of *S* and *R* will retard the establishment of the *S-R* tendency. On the other hand, in place learning, the stimuli of each arm are discriminative in a continuous fashion from the end box to the choice point area, i.e., there is an overlap of *CS* and *US*, and therefore faster learning (23) than under pure response conditions.

---

[7] Some learning of a nondifferential sort presumably does occur, and constitutes what is known as "cue familiarity" (11, 13).

Learning in the Skinner Box

Let us assume that we are attempting to train a rat to depress a bar in a conventional free-responding Skinner-box situation. A small pellet of food will be used as a reward and the food tray is located in the vicinity of the bar. *S* is given a minimal amount of food association to the food tray. Then the bar is introduced and thereafter food is delivered only following a full bar depression. Release of food is immediate and is accompanied by a "click."

The initial behavior of *S* in the box consists of exploratory behavior including operant bar depressions. However, the first definite response which appears to be learned is approach to the food tray. Precise quantitative data on the early learning of this approach response have been obtained by Hurwitz.[8] Initially then, as would be clearly predicted from our acquisition hypothesis, *S* simply learns to approach the food tray—both during the food association period and in the early stages of free responding to the bar. After this approach response has been established, *S* will approach the food tray without first having pressed the bar. Since on this occasion there will be no food in the tray with which food association has been established, the conditions for secondary elicitation exist. Consequently, all cues associated with *direct* approach to the food tray will acquire the property to elicit avoidant-type response. The responses elicited by the empty food tray include retreat from the tray and an increase in the number and vigor of tangential responses. With respect to bar pressing, this frequently results in a "flurry effect" or a high bar-press/approach-to-food-tray ratio. After one or more bar presses, *S* will eventually reapproach the food tray and procure the food pellet, for any avoidance to the tray has only been incipiently established. A variable which can facilitate the learning of approach to the food tray subsequent to pressing the bar is the response-produced click which has been previously associated with the presentation of food.

It is now appropriate to point out that *S* in a Skinner box is confronted with a discrimination situation. In part, this is to say that certain cues are associated with approach to food tray, and certain cues are associated with avoidance of the food tray. In addition, since it is likely that a flurry of bar presses has been elicited (reinforced) through secondary elicitation, it is necessary to explain why *S* learns to press the bar only once per visit to the food tray. The principal discriminative stimuli employed by *S* in learning to approach and press the bar *prior* to visiting the food tray are presented below in order of establishment: (*a*) Proprioceptive, visual, tactual, and auditory elements which occur immediately after pressing the bar; (*b*) those stimuli involved in addressing the bar; (*c*) those stimuli associated with orienting toward the bar.

The manner in which these cues mediate the chaining of the response elements of the bar-pressing sequence will now be discussed. To start with, "gross" or unconditioned manipulation of the bar results in unique stimulation which frequently involves a distinctive "click." These stimulus elements or their perseverative trace eventually occur contiguously with the response of approaching the food tray. Such stimuli therefore acquire the property to elicit approach to the food tray as soon as they occur. This means that *S* will learn to approach the food tray as soon as he presses the bar, i.e., when *S* hears the "click" or is otherwise stimulated by this distinctive stimulus complex. This also means that partial responses which do not activate the food release mechanism do not produce the stimuli which elicit approach to food tray.

---

[8] Personal communication from H. M. B. Hurwitz, Birkbeck College, London.

Given the establishment of this first major link in the behavior chain (approaching the food tray *after* pressing the bar), it becomes necessary for us to consider the chaining of the second link (pressing the bar *before* approach to the food tray) to the first link. This is straightforward. Since visual, tactual, and proprioceptive stimuli of the bar area occur in contiguity with bar pressing and approach to bar, they become conditioned to pressing and approaching the bar. Another factor which is essential to the chaining in of this response link is the fact that when *S* fails to press the bar or makes an incomplete bar-pressing response before visiting the food tray, avoidant-type responses to the food tray are elicited (reinforced). In short, cues associated with direct approach to the food tray and "short-cutting" responses acquire avoidance properties; thus these responses are extinguished. Therefore, through the interplay of primary and secondary elicitation the instrumental learning of a bar-pressing response turns out to be an elimination of one behavior sequence and the introduction and establishment of a more complex sequence. Or, as in the case of **T**-maze learning, instrumental bar depressing resolves itself into a case of discrimination learning.

Evidence in support of this discrimination-type analysis is available in a study from our laboratory. In this experiment, after bar pressing had been established to asymptotic level under continuous reinforcement, groups of *S*s were given varying amounts of 4:1 fixed-ratio reinforcement. A "click" accompanied only the to-be-rewarded bar press. Approach to the food tray following each of the first four bar presses in a fixed-ratio block tended to become extinguished. Following the fifth or to-be-rewarded bar press, *S* continued to visit the food tray. Presumably, the first four bar presses of a block tend to "discriminate out" and become chained in an instrumental fashion to the terminal or to-be-rewarded bar press. Given perfect discrimination, all responses of a block would be chained together and would constitute a single response unit. However, the end result of *any* amount of discrimination during the fixed-ratio schedule is to increase the number of individual bar presses which will occur during extinction. In the study being considered, it was found that resistance to extinction was *proportionate* to the level of discrimination attained.

SUMMARY

The present paper represents an attempt to explain two major types of instrumental learning within the framework of elicitation theory. Learning in a **T** maze was described as a case of discrimination learning, or a matter of learning to approach the cues on one side of the choice-point area and to avoid the cues of the other side. Pure response learning was contrasted with place learning.

Learning a bar-pressing response in a Skinner box situation was likewise described as a case of discrimination of a somewhat more complex nature. Learning to make a discrete bar press followed by an immediate approach to the food tray is a matter of the elimination (extinction) of a behavior sequence of directly approaching the food tray, and the establishment of a more complex sequence of first approaching and pressing the bar followed by an approach to the food tray. Experimental evidence in support of this interpretation was cited.

REFERENCES

1. ADELMAN, H. M., & MAATSCH, J. L. Resistance to extinction as a function of the type of response elicited by frustration. *J. exp. Psychol.* (in press).
2. AMSEL, A., & ROUSSEL, J. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.*, 1952, **43**, 363–368.
3. AMSEL, A., & WARD, J. S. Motivational properties of frustration: II. Frustration drive stimulus and frustration reduction in selective learning. *J. exp. Psychol.*, 1954, **48**, 37–48.
4. BARLOW, J. A. Secondary motivation through classical conditioning. *Amer. Psychologist*, 1952, **7**, 273. (Abstract)
5. DENNY, M. R., & DUNHAM, M. D. The effect of differential nonreinforcement of the incorrect response on the learning of the correct response in the simple T maze. *J. exp. Psychol.*, 1951, **41**, 382–389.
6 GLANZER, M. The role of stimulus satiation in spontaneous alternation. *J. exp. Psychol.*, 1953, **45**, 387–393.
7. GLANZER, M. Stimulation satiation: An explanation of spontaneous alternation and related phenomena. *Psychol. Rev.*, 1953, **60**, 257–269.
8. GOER, M. H. Position preference and discrimination learning. Unpublished doctor's dissertation. Michigan State Coll., 1954.
9. HARLOW, H. F. Motivation as a factor in the acquisition of new responses. In J. S. Brown *et al.*, *Current theory and research in motivation.* Lincoln, Nebraska: Univer. of Nebraska Press, 1953. Pp. 24–49.
10. HULL, C. L., & SPENCE, K. W. "Correction" versus "non-correction" method of trial-and-error learning in rats. *J. comp. Psychol.*, 1938, **25**, 127–145.
11. KARN, H. W., & PORTER, J. M., JR. The effect of certain pre-training procedures upon maze learning and their significance for the concept of latent learning. *J. exp. Psychol.*, 1946, **36**, 461–469.
12. LASHLEY, K. S., & WADE, M. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.
13. LAWRENCE, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39**, 770–784.
14. MAATSCH, J. L. Reinforcement and extinction phenomena. *Psychol. Rev.*, 1954, **61**, 111–118.
15. MONTGOMERY, K. C. A test of two explanations of spontaneous alternation. *J. comp. physiol. Psychol.*, 1952, **45**, 287–293.
16. NISSEN, H. W. The nature of drive as innate determinant of behavioral organization. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln, Nebraska: Univer. of Nebraska Press, 1954. Pp. 281–321.
17. NISSEN, H. W. Description of the learned response in discrimination behavior. *Psychol. Rev.*, 1950, **57**, 121–131.
18. PERKINS, C. C., JR. The relation between conditioned stimulus intensity and response strength. *J. exp. Psychol.*, 1953, **46**, 225–231.
19. RAZRAN, G. The conditioned evocation of attitudes (cognitive conditioning?). *J. exp. Psychol.*, 1954, **48**, 278–282.
20. SHEFFIELD, F., & ROBY, T. Reward value of a non-nutritive sweet taste. *J. comp. physiol. Psychol.*, 1950, **43**, 471–481.
21. SHEFFIELD, F., WULFF, J., & BACKER, R. Reward value of copulation without sex drive reduction. *J. comp. physiol. Psychol.*, 1951, **44**, 3–9.
22. SMITH, M. P., & BUCHANAN, G. Acquisition of secondary reward by cues associated with shock reduction. *J. exp. Psychol.*, 1954, **48**, 123–127.
23. TOLMAN, E. C., RITCHIE, B. F., & KALISH, D. Studies in spatial learning. II. Place learning versus response learning. *J. exp. Psychol.*, 1946, **36**, 221–229.

(Received for early publication February 11, 1955)

Psychological Review
Vol. 62, No. 4, 1955

THE PSYCHOLOGY OF MENTAL CONTENT RECONSIDERED[1]

DAVID C. McCLELLAND

Wesleyan University

Psychologists used to be interested in what went on in people's heads. In fact, for thousands of years this was practically all they were interested in. Psychologists from Aristotle to John Stuart Mill were concerned primarily with ideas and associations between ideas, but with the rise of modern scientific psychology we lost interest in ideas, by and large. The history of this development is well known, but let us review it for a moment. The psychology of mental content collapsed in the United States under the impact of two heavy blows. First, introspectionism seemed to run into a dead end. Titchener had argued manfully for a scientific study of the contents of mind, a kind of mental chemistry in which the basic elements would all be discovered and sorted out, but his laboratories simply failed to produce enough data to back up his theoretical position. It was not so much that his position was untenable. It was just that the data collected by the introspectionists did not seem to lead anywhere—to fruitful hypotheses, for example, which would serve to make theoretical sense out of the flux of mental events.

The second blow was even more devastating. It was, of course, the behavioristic revolution. Particularly in the United States, psychologists began to argue that conscious content could never form the basis of a science, whereas behavior could. J. B. Watson led the revolt in the name of scientific objectivity. After all, could you see or touch or feel or record with a machine a thought or a feeling? Now, a muscular contraction—an eye blink, a foot withdrawal, a right turn in a maze—that was something else again. That could be seen and touched and felt and often recorded entirely automatically by an impersonal, mechanical gadget. Here was the stuff of which a real science could be made!

Looking back with the perspective of 30 years we can begin to see why this movement was so appealing. In the first place, it did provide the kind of objectivity, methodologically speaking, that psychology had never had before and it could, therefore, lend real support to psychology's claim that it was a science. Secondly, it fit in with the traditional American pragmatic bias in favor of action rather than thought or feeling which were generally considered to be old-fashioned European concepts. After all, in the United States it is what a man does that counts, not what he says or thinks or feels. This tendency in American psychology is still so prevalent that to many of us prediction of behavior means *only* predicting gross motor behavior rather than predicting thoughts, conflicts, doubts, imaginings, feelings, etc. as reflected in verbal behavior. Thirdly, behaviorism tended to focus attention on problems which were of vital topical interest to a new country in which many of its citizens were attempting to adjust to new ways of life. In fact, adjustment or *learning* became the key concept. And this was natural in a country in which so many immigrants or their children had to give up traditional European ways of behaving for new American habit patterns. It was at this point that psychology became almost exclusively interested in "process variables," in how people went about doing things rather than in what they did. This was the time when Woodworth was stressing that we should rewrite psychology in terms of "ing" words—e.g., perceiving, emoting, thinking, learning, etc. No one was interested in *what* people thought, *what* they perceived, *what* they learned, etc. Instead we were to be concerned only with the laws which governed the *process* of perceiving, learning, etc.

---

[1] This paper was delivered at the Fourteenth International Congress of Psychology in Montreal, June 9, 1954.

Even personality and social psychology, which by definition are content oriented and which, therefore, should have resisted this trend, fell under the spell of this widespread movement. In personality psychology we were primarily interested in self-descriptive inventories in which the subject answered a lot of questions about his aches, pains, and anxieties. But, mind you, we did not look at his answers. We added them up to get a neuroticism score or dominance score or what not. We were not interested even here in *what* he said about himself, in *what* his ideas were. We were only interested in the extent to which his answer contributed to a total score which meant something else. To be sure, an individual clinician sometimes went so far as to look at the actual answers a person had given on a personality inventory, but then, he was not a scientist! In social psychology, too, we managed to get along without much concern with content, although here, too, it was a little difficult. The problem was solved with the help of the attitude concept. An attitude is essentially what I have been calling a process variable. We are interested in *how* attitudes get set up, in how we can measure them, in their consistency, in their rigidity, their generality or specificity, etc.—all process variables. But we are not interested in *what* they are, particularly. Any old attitude will do for our purpose, just as in studying learning, any old task will do for our purpose—a maze, a bar to press, a list of nonsense syllables, or what not. So social psychologists chose as the attitudes to be investigated whatever happened to be of current interest at the moment—e.g., internationalism, feminism, pacifism, race prejudice, and so on. Few if any people thought it was even worth asking which attitudes were the "important" ones to use in describing a person or his culture. Many people probably would have wondered whether such questions really fell within the province of psychology at all.

It is against the background of this widespread social movement in psychology that we can see the beginnings of the projective testing movement as the source of a change in attitude which is finally beginning to be felt today, possibly in large part because of the success of projective tests. But certainly projective testing did not start as a conscious revolt against this interest in process. Quite the contrary. The Rorschach test, as one of the oldest of these new instruments, probably gained as wide acceptance as it did in the United States largely because it became primarily process-oriented. It became concerned with *how* people perceive and only secondarily with *what* they perceive. Quantitative indices could be computed according to how many responses were determined primarily by form, by color, by movement, and the like. Nevertheless, the good clinicians often found that the particular content of the association given by the patient was of value to them in understanding the person. And this has always been the case with a good clinician. He *has* to be interested in *what* his patient thinks as well as in how he thinks it. Even though his formal psychological training gives him very little assistance at this level, he knows that in order to handle this particular person he has to be interested in the patient's ideas. This, it seems to me, has been the great and continuing contribution of the clinic and the projective test in a time when psychological theorists have talked themselves out of being interested in content altogether. I am reminded here of a comment made to me once by one of my more cynical colleagues who claimed that no new ideas of importance ever appeared in the universities. Usually they appear outside first, and then are only gradually claimed by the universities. Certainly if we think for a minute of men like Descartes, Darwin, Freud, or Einstein, there would seem to be something in what he said. And this development seems to be a case in point. The projective testing movement grew up largely *outside* the conservative academic tradition and finally, because of its clinical success, has managed to dent the calm assurance with which many theoretical psychologists have discarded all problems of mental content.

But to continue with my story: The real change came with the development by Murray and his associates of the Thematic Apperception Test about 20 years ago. Now for the first time we had an instrument in which the primary concern was not form but content. The person interpreting a TAT record must ask such questions as: *What* motives activate this person? *What* conflicts plague him? *What* modes of defense does he adopt? *What* characteristics does the world have for him? No longer are we concerned primarily, as in the Rorschach, with *how* he approaches his task, although some have attempted to analyze the TAT in these terms. To help us in our analyses of such content, we have drawn heavily on psychoanalysis, the one system of content psychology which, isolated in the clinic, survived the mass attack of behaviorism in the laboratory. Murray in his original system of analysis for the TAT attempted to provide us with a much broader vocabulary for the analysis of content, but, by and large, in our analyses we draw upon relatively few general psychoanalytical concepts such as sex, aggression, parent-child relationships, and the like. This, to my way of thinking, is an impoverished set of concepts for dealing with mental content, but it is nonetheless the one real and vital one in the United States today.

What evidence is there that this tendency to concern oneself with mental content is of growing influence? In the first place, we must not underestimate the conservative resistance to the belief that such a psychology is possible. Even Freud's generalizations about the importance of certain basic conflicts such as those involved in the Oedipus complex are under constant attack. To some extent the attacks are motivated by the conviction that generalizations about content are really impossible. The argument runs that there are no general concepts which can serve to describe the human situation at *any* place or time in history. What about cultural relativism? After all, individuals differ widely in what they think and so do cultures. Some have an Oedipus complex and some do not. How can we generalize about anything except the process by which individuals arrive at their ideas? The ideas themselves are completely relative. One can perhaps be literary about them, but not scientific. This is the argument and there is no answer to it, except to prove that it can be done fruitfully. Many of us are convinced already, for example, that despite individual and cultural variations, it is a major scientific achievement to have focused attention on the framework of the mother-son relationship as of primary importance in the development of the individual, and to have worked out some of the taxonomy of this and the allied relationship with the father.

Meanwhile, there have been some new developments which would encourage us to believe that perhaps a psychology of content is possible. Take the research report on *The Authoritarian Personality* (1), for example. I would contend that the essential issues raised by this research are issues in the psychology of content. It represents to some extent a fusion of psychoanalytic structural concepts with certain concepts drawn from political ideology. Whether we like the fusion or not does not really matter too much. From the methodological point of view it represents an exciting step forward since its authors have drawn on political ideology as well as psychoanalysis to help explain the structure of personality. When our science has matured to the point where we can draw not only on political ideology, but on economic, religious, esthetic ideologies, and the like, then we will be on the way toward developing a really full-blown psychology of mental content.

Our own research on *The Achievement Motive* (2) has contributed as much to these conclusions as anything else. We started with the relatively simple task of identifying those types of imagery in a TAT-type record which indicated the presence of a motive to achieve or succeed. After we were able to identify reliably this item of mental content, we were able to select individuals whose thought processes contained a lot of such items and other individuals whose thought processes contained few such items. We were then faced with the question of how these people differed. Do they behave differently? Yes, they do. The ones with a lot of achievement imagery tend to learn faster, to perform better, to set different levels of aspiration, to have a better memory for incompleted tasks, to perceive the world in different terms, etc. Perhaps even more interesting was the question of how they got that way. How is it that some people tend to think more often in achievement terms? We were led back to the mother-son relationship and found that independence training seems to be associated with achievement motivation. That is, those mothers who encouraged their sons to develop independently, to learn their way around by themselves, seemed to have sons with higher achievement motives. But we pushed the question one step further back. How is it that some mothers favor independence training more than others? This raised the question of values, and values raised the question of religious ideology. Then we found that attitudes toward independence training were not randomly distributed through various population subgroups. Instead, Protestant and Jewish parents were much more likely to favor early independence training than were Catholic parents and this, in turn, seemed to fit into the belief systems and emphases of these three religions (3). And if this is so, we can begin to trace some of the details of the connection between Protestantism and the rise of capitalism as originally outlined by Max Weber (5) and R. H. Tawney (4).

So our recent research has led us into the relationships between religious values, independence training, achievement motivation, and economic development. We think we are beginning to discover some connections among these phenomena which can be traced out with a fair degree of scientific confidence. But whether we succeed or not, the point I am trying to make is that by concentrating on one item of mental content, namely achievement imagery, we have opened up a whole new set of problems in social science that can be investigated by psychology.

Now let us pause a minute and try to reconsider what has happened. The "new look" in the study of mental content really involves neither introspection nor projective testing in their pure forms. Instead I prefer to call it "thought sampling," and to use the analogy of the "blood count" from the medical laboratory to explain what I have in mind. Just as we need a sample of blood to make a white cell count, we want a sample of thoughts or ideas to make our imagery counts. In general, we get these samples by asking the subject to write stories to certain cues, usually verbal or visual. Having gotten our thought sample, we have to learn to recognize certain types of imagery, just as the medical technician has to learn to recognize a white blood cell when he sees one. This involves a great deal of preliminary work so that we can define the characteristics of the imagery carefully and then train individuals to recognize what they are looking for. It does require training, but probably no more training than a medical technician needs to be able to distinguish one type of blood cell from another. That is, it does not involve high-level judgment, but is essentially a "pointing" operation which is a little, but not much, more difficult to make than pointing to an animal's right turn in a maze. Watson need not have feared for the objectivity of this kind of analysis. The record is permanent. The same person can look at it again and again, or several people can analyze it. It is quite possible to get a reliable and objective result.

If we put the "new look" in mental content in these general terms, it is immediately clear that we have a number of problems to solve. For example, there is the sampling problem. Under what conditions should we get our thought samples? What cues should we use? Should we get long samples or short samples? What about the subject's set? How does it influence content? Here it becomes obvious that traditional projective testing elicits only a very small segment of the possible types of mental content. To the extent that we stay within the limited framework of the traditional TAT cards, for example, we are bound to have a biased sample of what goes on in a man's mind.

An even more important problem has to do with the decision as to what categories for content analysis we use. This is the heart of the problem, since we will get theoretically meaningful results, or generalizations that hold for a wide variety of situations, only if we choose the right categories to begin with. How does one discover the right categories? The literature of social science is strewn with content analyses of everything from open-ended interviews to "soap operas"—analyses which are purely *ad hoc*—for the immediate practical purpose in hand. I am certainly not arguing for more of this industrious busy work. The categories must be meaningful; they must be related to theory; they must be trans-situational—i.e., applicable to more situations than the one to which they are first applied. It takes inspiration or luck or hard work or something to discover such a category, just as it did in biology to discover what was the most useful of many possible ways to classify blood cells. The only concrete suggestion I have as to how to proceed, which comes from our own experience, is to choose those categories which show significant shifts as a result of experimental operations. Whether this is an unnecessarily restrictive rule, I do not know. At any rate it certainly eliminates many possible categories, and it seems to be roughly the one which the chemists have used in setting up the classification of elements.

This brings us back to Titchener. Looking back with the perspective of history, we can now see that Titchener's structural psychology failed for two reasons. In the first place, the content categories he chose did not turn out to be fruitful. They were not related to experimental operations on the one hand or to other types of behavior on the other. For these or other reasons, they simply did not lead to theoretical development. Therefore they were the wrong categories. In the second place, and this is of major methodological importance, his students categorized their own data. The essence of introspection is that the same person serves both as a source of data and as a categorizer of them. This has an obvious weakness, a weakness which has been perpetuated in self-descriptive personality inventories. It is simply that the subject may have a very imperfect or incorrect idea of what categories his thoughts belong in. This may be because his categories are different from the ones we as scientists want to use or because he may really misperceive himself. The great contribution that both Freud and the projective testing movement made was that neither asked the subject to pass judgment on his thoughts as they appeared to him. Both simply asked for a sample of those thoughts and then left the categorization process to an outside observer. This was an important methodological advance, the significance of which I think we are only just beginning to appreciate.

If psychologists are to re-enter the field of mental content and start classifying it according to categories of genuine theoretical fruitfulness, I fear they will have to return to disciplines they have long neglected. In the twenties, in the heyday of behaviorism, we were proud that we knew nothing of religion, of art, of history, of economics, or politics (except in a personal, often naive way). We didn't need to know about these things if we were only interested in process variables. We could make our own choice of a task situation—for example, the rat in a maze—and what we found out there about how the rat learned the maze would apply equally well to *all* (including human) learning situations. We could afford to be ignorant of many things that man has thought about. But if the psychology of content develops as I think it will, we shall have to go back to getting a broad, general education. Certainly nothing in my training *as a psychologist* prepared me to handle problems in religious belief systems or economic development. Yet these are typical of the problems which I think will begin to arise increasingly often in the new psychology of content, and we simply cannot afford to be naive and pretend that research scholars in these fields have nothing to tell us.

If my analysis is correct, we are on the brink of an important new development in psychology. Because of methodological improvements, we are about to take up again some of the problems in mental content that formerly were considered to be an essential part of psychology. And it is my conviction that the projective testing movement is to be thanked for keeping an interest in content alive in an era when most theoretical psychologists were otherwise occupied, and for providing us with the methodological advance that enabled us to escape from the blind alley into which introspection had led us.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.
2. McCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
3. McCLELLAND, D. C., RINDLISBACHER, A., & DECHARMS, R. Religious and other sources of attitudes toward independence training. In D. C. McClelland (Ed.), *Studies in motivation.* New York: Appleton-Century-Crofts, 1955. Pp. 389–397.
4. TAWNEY, R. H. *Religion and the rise of capitalism.* New York: Harcourt, 1926.
5. WEBER, M. *The Protestant ethic.* (Trans. by Talcott Parsons.) New York: Scribner's, 1930.

(Received October 1, 1954)

Psychological Review
Vol. 62, No. 4, 1955

ON MAKING PREDICTIONS FROM HULL'S THEORY

JOHN W. COTTON[1]
Northwestern University

The theory of behavior proposed by Hull in *Principles of Behavior* (5) and modified several times thereafter (6, 7, 8) has called forth numerous experimental studies and theoretical articles focusing upon the empirical plausibility of its postulates. The present paper is not concerned with this question of plausibility. Rather, we wish to inquire whether the theory as stated leads logically to predictions which may be tested empirically and, if so, whether these predictions include any sizable proportion of the theorems commonly supposed to follow from the theory. We will conclude that the answer to the first question must be "No," except under special circumstances, and that the answer to the second question is also "No." These answers will result from application of a strict interpretation of the theory, an interpretation which may seem unduly restrictive to some theorists. It should be noted, however, that only a highly developed theory such as that of Hull could be subjected to an analysis of the type to be presented. The reader must judge for himself whether the conclusions reached in this inquiry are reasonable.

Our remarks will have primary reference to *Principles of Behavior* because it is the most widely known presentation of the theory. The conclusions reached would also result from consideration of any of the revised postulate systems which Hull proposed.

The first impression of a person reading Hull's system or the body of literature surrounding it might well be that a vast number of predictions from the theory have been made and tested. Because a considerable amount of data was presented and discussed in Hull's formulation of his theory, it is tempting to conclude that this material at least has already been shown to be consistent with the theory. In fact no one has publicly questioned the presumption that the empirical data cited in *Principles of Behavior* is embraced by the theory. Yet this presumption is false. There was, to be sure, a romance between theory and fact, but the wedding did not take place. Hull never quite compared the implications of his theory with the data he discussed.

AN EXAMPLE OF THE PREDICTIVE DEFICIENCIES OF HULL'S THEORY

To illustrate the respects in which Hull's theory fails to make predictions consistent with the data presented when the theory was first formulated, let us consider a particular experiment. This experiment, the first discussed in Hull's development of the construct called *habit strength* ($_SH_R$), is Hovland's (4) study of the relationship between the number of reinforced trials (N) and the amplitude (A) of the galvanic skin reaction. Hull summarized Hovland's results with the following equation, which is (except for notation) the former's equation 4:

$$\mu_A = 14.1(1 - 10^{-.033N}) + 3.1 \qquad [1]$$

---

[1] The author expresses appreciation to his colleagues, D. T. Campbell, C. P. Duncan, W. A. Hunt, D. J. Lewis, and B. J. Underwood, for a critical reading of this paper and for several suggestions which were incorporated in the paper in its present form.

where μ_A is the population mean amplitude of response estimated from Hovland's data. Equation 1, then, is a statement of an empirical law which Hull's theory might reasonably be expected to encompass. We will require of the theory that it yield equation 1 as a theorem, or at least that it yield a theorem consistent with the data upon which this equation was based. Since mathematical statements of the postulates are presented by Hull, it will be possible to make mathematical deductions in order to permit comparison of equation 1 with a theorem stated in mathematical form.

Finding the value of ${}_SH_R$. The first step in the generation of a theorem relating to μ_A to N is to tie N, an antecedent condition, to Hull's intervening variables. Then the intervening variables may be tied to μ_A, a consequent condition, and the prediction will be complete. Accordingly, we now consider ${}_SH_R$, the first intervening variable involved in the predictive chain. On the basis of the Hovland experiment and others, Hull defined ${}_SH_R$ by his equation 16, the mathematical statement of his Postulate 4. We present an equivalent definition, Hull's equation 26:

$${}_SH_R = m(1 - 10^{-iN}) \qquad [2]$$

where m is the maximum habit strength obtainable under any given experimental situation and i is a constant reflecting the rate of approach of ${}_SH_R$ to m. Hull's equation 26 is used here in the interest of simplicity: his equation 16 has three factors whose values depend upon experimental variables held constant by Hovland. Therefore all three factors may be subsumed under a constant value of m.

Finding the value of ${}_S\bar{H}_R$. The second intervening variable to be considered is *effective habit strength* (${}_S\bar{H}_R$). Hull's equation 29, the mathematical statement of his Postulate 5, may be combined with our equation 2 to yield the following expression:

$${}_S\bar{H}_R = m(1 - 10^{-iN}) \qquad [3]$$

if we presume that no difference existed in stimulation at the time of training and the time of testing in Hovland's experiment.

Finding the value of ${}_{S_1+S_D}\bar{H}_R$. Now it becomes necessary to relate N to the total effective habit strength (${}_{S_1+S_D}\bar{H}_R$), by combining ${}_S\bar{H}_R$ with the effective habit strength loading of the drive stimulus (${}_{S_D}\bar{H}_R$). First we infer that ${}_{S_D}\bar{H}_R = {}_S\bar{H}_R$ on the basis of a comparison of columns 3 and 8 of Table 5 of *Principles of Behavior*. Then we employ our equation 3 together with Hull's equation 30, the mathematical statement of Major Corollary I, in the special form given by the seventh equation on page 255, to establish the relationship between ${}_{S_1+S_D}\bar{H}_R$ and N:

$${}_{S_1+S_D}\bar{H}_R = m[1 - 10^{-iN}] \times \left[2 - \frac{m(1 - 10^{-iN})}{100}\right]. \qquad [4]$$

Finding the value of ${}_SE_R$. The process of deduction must now be continued with the specification of the manner in which reaction potentiality (${}_SE_R$) depends upon N. Using Hull's equation 34, the mathematical statement of Major Corollary II, and our equation 4, we obtain the required equation:

$${}_SE_R = m[1 - 10^{-iN}] \times \left[2 - \frac{m(1 - 10^{-iN})}{100}\right] D' \qquad [5]$$

where D' is our simplification for the expression $(\dot{D} + D)/(\dot{D} + M_D)$, which is a constant for the Hovland experiment. D', then, represents the combined magnitude of all drive effects in Hovland's experiment.

Finding the value of $\dot{I}_R$. The development of a predictive equation relating μ_A and N now involves the determination of the relationship between the amount of net inhibition ($\dot{I}_R$) and N. At this point Hull's Postulates 8 and 9 must be considered. We have no choice except to conclude that these postulates, as stated, provide no basis for a specification of the relationship desired. Although Postulate 9 states a rule for summating reactive inhibition (I_R) and conditioned inhibition (${}_SI_R$) to obtain $\dot{I}_R$, neither postulate gives much assistance in determining the values of I_R and ${}_SI_R$.

The logical errors of Postulates 8 and 9 are as follows: (*a*) Equation 36, the mathematical statement of Postulate 8a, contradicts equation 42, the mathematical statement of Postulate 9a, except when ${}_SI_R = 0$. (*b*) In equation 37, the mathematical statement of Postulate 8b, the symbol n is used to represent the number of reaction evocations producing inhibition. Since n is employed on the same page to represent the number of unreinforced reaction evocations required to produce extinction, this usage contradicts the verbal statement in Postulate 8 which asserts that, *whenever* a reaction is evoked, inhibition will be created. (*c*) The verbal Postulates 8b and 8c refer to $\dot{I}_R$, whereas the mathematical statements of those postulates refer to ${}_SI_R$. (*d*) If Postulates 8b and 8c are taken to refer to ${}_SI_R$ in order to provide an independent definition of one component of $\dot{I}_R$, then I_R is still left without a definition, and vice versa. (*e*) The symbol I_R in equation 41, the mathematical statement of Postulate 9a, must be replaced by $t'''I_R$, the symbol for the amount of I_R remaining after inhibition has been allowed to dissipate for time t'''. Otherwise equation 40, the mathematical statement of Postulate 9d, can play no part in the prediction of behavior under any circumstance to which the theory may be applied.

Most of the inconsistencies and omissions just noted have previously been discussed either by Koch (**10**) or by Montgomery (**12**). These inadequacies of the theory may be removed by rewriting and expanding Postulates 8 and 9. Several possible modifications could be proposed, either on theoretical or empirical grounds. However, we present the revisions which seem most in keeping with Hull's intentions when writing these postulates. Only the mathematical statements of the revised postulates need be given here:

Revised Postulate 8a:

$$ {}_S\bar{E}_R = {}_SE_R - \dot{I}_R. \qquad [6] $$

Revised Postulates 8b and 8c:

$$ I_R = (cN')/(B - W). \qquad [7] $$

Revised Postulate 9b:

$$ \dot{I}_R = t'''I_R + {}_SI_R - \frac{(t'''I_R)\ ({}_SI_R)}{100}. \qquad [8] $$

Postulate 9c:

$$ {}_SI_R = I_R. \qquad [9] $$

The above postulates are for the most part self-explanatory. It will be noted that we have only one equation for Postulates 8b and 8c, and that N' is used in place of n. Only one equation is necessary because Hull's mathematical statements of Postulates 8b and 8c are algebraically equivalent. The symbol N' should be taken to refer to the number of reaction evocations observed either during acquisition or extinction or a habit, or during both acquisition and extinction. In Hovland's experiment

$N' = N$ because testing followed N reaction evocations.

Since Postulates 8d and 9b require no modification, they have not been presented above. Postulate 9c will be seen to be completely new to the system. We have no particular faith in its adequacy, but introduce it only because some postulate about ${}_SI_R$ is essential if predictions are to be made within a theory containing Postulate 9a.

The modifications of Postulates 8 and 9 proposed above permit us to derive an equation relating $\dot{I}_R$ to N, as was our original purpose in this section. Since t''' was constant in Hovland's experiment, we replace the term $e^{-qt'''}$ by the constant k and write the following equation based upon Postulates 8 and 9 in their revised form:

$$\dot{I}_R = \frac{cN}{(B-W)} \times \left[1 + k - \frac{ckN}{100(B-W)}\right]. \quad [10]$$

Finding the value of ${}_S\bar{E}_R$. A specification of the relationship between Hull's next construct, effective reaction potential (${}_S\bar{E}_R$), and N may be obtained by substituting the quantities from the right-hand sides of equations 5 and 10 into equation 6 to yield the expression:

$$\begin{aligned} {}_S\bar{E}_R = {} & m[1 - 10^{-iN}] \\ & \times \left[2 - \frac{m(1 - 10^{-iN})}{100}\right] D' \\ & - \frac{cN}{(B-W)} \\ & \times \left[1 + k - \frac{ckN}{(100)(B-W)}\right]. \quad [11] \end{aligned}$$

Finding the value of ${}_S\dot{\bar{E}}_R$. One last intervening variable remains to be defined in terms of N before a prediction of amplitude as a function of N becomes possible. This construct, *momentary effective reaction potential* (${}_S\dot{\bar{E}}_R$), is defined by Hull's equation 44, the mathematical statement of his Postulate 10:

$${}_S\dot{\bar{E}}_R = {}_S\bar{E}_R - {}_SO'_R \quad [12]$$

where ${}_SO'_R = {}_SO_R$ when ${}_S\bar{E}_R \geq {}_SO_R$ and where ${}_SO'_R = {}_S\bar{E}_R$ when ${}_S\bar{E}_R < {}_SO_R$, the range of ${}_SO_R$ being specified by the relationship $0 \leq {}_SO_R \leq 6\sigma$. The distribution function of ${}_SO_R$ is that of a normal variable truncated three σ's above and three σ's below its mean of 3σ:

$$f({}_SO_R) = \frac{1}{.9973\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{{}_SO_R - 3\sigma}{\sigma}\right)^2}. \quad [13]$$

Substituting the value of ${}_S\bar{E}_R$ from equation 11 into equation 12, we obtain the desired relation between ${}_S\dot{\bar{E}}_R$ and N:

$$\begin{aligned} {}_S\dot{\bar{E}}_R = {} & m[1 - 10^{-iN}] \\ & \times \left[2 - \frac{m(1 - 10^{-iN})}{100}\right] D' \\ & - \frac{cN}{(B-W)} \\ & \times \left[1+k-\frac{ckN}{100(B-W)}\right] - {}_SO'_R. \quad [14] \end{aligned}$$

Predicting A from N. Given equation 14, it is now possible to use one final postulate of Hull's system to predict the value of the response measure A as a function of N. Substituting from equation 14 into Hull's equation 50, the mathematical statement of his Postulate 15, we obtain:

$$\begin{aligned} A = h' \Big\{ & m[1 - 10^{-iN}] \\ & \times \left[2 - \frac{m(1 - 10^{-iN})}{100}\right] D' \\ & - \frac{cN}{(B-W)} \Big[1 + k \\ & - \frac{ckN}{100(B-W)}\Big] - {}_SO'_R \Big\} - i'. \quad [15] \end{aligned}$$

Equation 15 would be completely acceptable for predictive use if it were not that ${}_sO'_R$ is a random variable. Consequently A can be estimated from equation 15 once the parameters of that equation are known, but the error of prediction will sometimes be as large as the range of possible ${}_sO'_R$ values, i.e., up to six σ's. For this reason we would customarily not attempt to predict A. Rather we turn our attention to the prediction of a measure of central tendency of the A distribution.

Predicting μ_A *from* N. All the common measures of central tendency could be determined for the A distribution. Since the Hovland experiment was a study of the mean alone, we consider only μ_A. Two problems arise at this point: (*a*) Since Hovland employed several different subjects in his experiment, we must ask whether the different persons should be presumed to have different values for the constants in equation 15. (*b*) We must determine the mean of an ${}_sO'_R$ distribution which is always truncated and in addition sometimes has a discrete probability density at the point ${}_sO'_R = {}_s\bar{E}_R$.

The first problem raised is so complex that it must be treated in detail later in this paper. In the interests of simplicity, we assume for the present that all subjects given N reinforced trials will have identical ${}_s\bar{E}_R$ values, identical ${}_sO'_R$ distributions, identical h' values, and identical i' values. In this case the population mean amplitude of response, μ_A, is equal to h' times the population mean of the bracketed term in equation 15 less i'. Since the only random variable in this bracketed term is ${}_sO'_R$, the determination of μ_A depends primarily upon the solution of problem *b* above.

We present without proof the equations for $\mu_{{}_sO'_R}$:

$$\mu_{{}_sO'_R} = 3\sigma \quad \text{when} \quad {}_s\bar{E}_R \geq 6\sigma \qquad [16a]$$

and

$$\mu_{{}_sO'_R} = {}_s\bar{E}_R \int_{{}_s\bar{E}_R}^{6\sigma} f({}_sO_R)\,d{}_sO_R + \int_0^{{}_s\bar{E}_R} {}_sO_R\, f({}_sO_R)\, d{}_sO_R \quad \text{when} \quad {}_s\bar{E}_R < 6\sigma. \qquad [16b]$$

On the basis of equations 16a and 16b, the bracketed term of equation 15 has a known mean, and the value of μ_A becomes:

$$\mu_A = h'\left\{ m[1-10^{-iN}] \times \left[2 - \frac{m(1-10^{-iN})}{100}\right] D' - \frac{cN}{(B-W)}\left[1+k-\frac{ckN}{100(B-W)}\right] - 3\sigma \right\} - i' \qquad [17a]$$

$$\text{when} \quad {}_s\bar{E}_R \geq 6\sigma$$

and

$$\mu_A = h'\left\{ m[1-10^{-iN}] \times \left[2 - \frac{m(1-10^{-iN})}{100}\right] D' - \frac{cN}{(B-W)}\left[1+k-\frac{ckN}{100(B-W)}\right] - {}_s\bar{E}_R \int_{{}_s\bar{E}_R}^{6\sigma} f({}_sO_R)\, d{}_sO_R - \int_0^{{}_s\bar{E}_R} {}_sO_R\, f({}_sO_R)\, d{}_sO_R \right\} - i'$$

$$\text{when} \quad {}_s\bar{E}_R < 6\sigma. \qquad [17b]$$

How the prediction and the data compare. Equations 17a and 17b are the culmination of the process of predicting the functional relationship between μ_A and N from Hull's postulates and a consideration of the experimental design employed by Hovland. Equation 17a can be compressed into an expression suitable for curve fitting, and will then contain six constants. It will have terms involving N to the first and second degree as well as terms involving 10^{-iN} and 10^{-2iN}. Equation 17b cannot be employed at all unless a special assumption is made regarding the value of ${}_S\bar{E}_R$ relative to σ at some value of N. Furthermore, one cannot tell when to use equation 17a in preference to equation 17b until this assumption is made. Thus we know little more than that predictions could be made with equation 17a if N were greater than some unknown value.

Under the presumption that this required N is quite small, we compare equation 17a with Hovland's findings as a means of testing the predictive utility of Hull's theory. The result of this comparison is not favorable: equation 1, the result of Hull's curve fitting of the empirical relationship observed by Hovland, bears almost no resemblance to equation 17a. Since equation 1 cannot be generated from the theory, it must be considered an empirical equation only.

From the above we conclude that the theory does not meet our requirement that it should yield equation 1 as a theorem. The possibility remains that equation 17a is consistent with the data upon which equation 1 was based rather than with equation 1 itself. However, since Hovland obtained data for only five values of N whereas equation 17a contains six constants, a fitting of his data to that equation would be a trivial accomplishment.

The predictive process just completed need not be taken as conclusive evidence that Hull's theory contradicts any data from which he built that theory. By no means, however, should it be taken as confirming evidence for the theory. If equations 17a and 17b could be fitted to experimental data having more values of N than Hovland's data, the prediction might be considered adequate. Even then, however, the theory as originally stated would be considered objectionable for several reasons: (*a*) Revision of Postulates 8 and 9 was necessary before any prediction at all could be made. (*b*) The restrictions imposed upon ${}_SO'_R$ at the value ${}_SO_R > {}_S\bar{E}_R$ cause the derivative of the μ_A function to be discontinuous when ${}_S\bar{E}_R = 6\sigma$, a prediction which is conceivable but unattractive from an empirical point of view. (*c*) Special assumptions will have to be made regarding the constants in equation 17a before curve fitting with it can be attempted. (*d*) Restrictions will have to be placed upon the constants in equation 17b if it is to be a monotonic increasing function of N, as one would expect from existing data.

The foregoing remarks have been designed to show the implications of Hull's theory as it relates to a specific experimental study. Comparable conclusions would be reached in the theoretical analysis of any other study discussed in *Principles of Behavior*. In all cases the predictive equations derived from the theory as stated will fail to conform to Hull's empirical equations. The predictive equations will be much too complex, a result of undue complexity in the definition of ${}_{S_1+S_D}\bar{H}_R$, $\dot{I}_R$, and ${}_SO_R$. These difficulties could be eliminated by appropriate modification of the theory. This modification will not be attempted here, however, because our present purpose is to consider the properties of the theory as it now exists.

DEGREES OF SPECIFICITY IN MAKING PREDICTIONS FROM A THEORY

The theory in question might be defended by some of its advocates on the ground that we have taken its postulates too seriously—in short, that it should not be considered a system from which predictions as explicit as those of equations 17a and 17b should be made. The argument, we are sure, would not have been employed by Hull, since he employed almost all conceivable means of mathematizing with his theory and his data, except the approach presented above. It is true, however, that many of Hull's corollaries are stated in qualitative form and that many experiments relating to the theory have been designed to test predictions less involved than the one just stated. We now inquire whether the occurrence of a verification or verifications of predictions of a simple nature can mitigate the dissatisfaction produced by the results presented above, and how predictions of this nature should be made.

The prediction of behavior to be expected in a given experimental situation may be as coarse as predicting from an inequality in magnitude of some experimental variable to a corresponding inequality in some response measure. It may, on the other hand, be as fine as the prediction of the numerical value of some parameter in an equation relating the experimental and response variables. The range of coarseness of predictions which have been made from Hull's theory is illustrated by three experiments, performed by Meehl and MacCorquodale (**11**), Siegel (**14**), and Burke (**1**).

The prediction of inequalities. In an attempt to explain latent learning phenomena, Meehl and MacCorquodale (**11**) state that feeding animals in a box, outside a six-unit Blodgett multiple-T maze to be learned, should arouse a secondary drive state which would increase the over-all magnitude of drive ($\bar{D}$). This, in turn, would be accompanied by an increase in $_{S}\bar{E}_{R}$ for animals thus fed as compared with animals not given this feeding. The differential $_{S}\bar{E}_{R}$ values for these two groups of animals are then taken to imply that the animals with the greater $\bar{D}$ will show shorter mean running times through the maze, smaller mean Blodgett errors, and a smaller sum of weighted errors than the other group. Each of these predictions was verified experimentally by Meehl and MacCorquodale, thereby supporting their notion that latent learning may be explained in some cases as the result of a modification in drive at the time reward is introduced for the first time.

The prediction of the general form of a relationship. A second level of specificity in prediction is exhibited in the report of an experiment by Siegel (**14**). This experimenter forced human subjects to make from 0 to 160 pressing responses with a switch on their right side before permitting 10 free choice responses in which either a left or a right switch could be selected on each trial. Using the verbal statement of Hull's Postulate 8b, Siegel stated that the total amount of inhibition ($\dot{I}_R$) would be a linear function of the number of forced choices (N'). He then predicted that the mean of the proportions (p') of left responses would also be a linear function of N'. Siegel's predictions were verified, thus giving apparent support to Hull's theory.

The prediction of numerical values of parameters. The prediction of exact numerical values of parameters in theoretical equations may be based almost completely upon theoretical considerations, or it may be based upon the determination of its value from previous data, in which case the experimenter attempts to recover that value with new data. The latter method only will be considered here.

An example of this type of prediction is found in a study by Burke (1), who worked with a somewhat modified version of Hull's theory. For each animal used in his experiment, Burke first fitted a theoretical cumulative distribution function of the asymptotic running times obtained in a straight runway. This provided a test of the theory at the second level of specificity, the comparison of empirical data with equations having a given general form. When the theory proved satisfactory at this level, Burke proceeded to substitute the exact values of parameters from these asymptotic equations into the theoretical equations for preasymptotic trials. Since the latter equations were developed independently of the preasymptotic data which they were designed to describe, this prediction of the magnitude of preasymptotic running times could be compared with the running times actually obtained, to provide a test of the predictive usefulness of the exact numerical parameter values. Because the theoretical equations described the preasymptotic data satisfactorily, the constants taken from the asymptotic data were said to be recovered in the preasymptotic data.

The relative merits of the three types of prediction. With examples of the three levels of specificity before us, we may inquire as to the advantages possessed by each method. First, we will consider the amount of deductive effort required, and then we will ask which methods provide the most severe tests of the theory.

Because a prediction from an inequality in an antecedent variable to an inequality in a consequent variable is seldom accompanied by a long demonstration that the first inequality implies the second, it may appear that this type of prediction involves less deductive labor than do the other types of prediction. In some cases this difference may in fact be a legitimate reflection of a reduction of effort required to make a prediction. In most cases, however, Hull and the other workers making predictions of inequalities have not actually generated their theorems or corollaries by strict application of logical operations to Hull's postulates. Thus Meehl and MacCorquodale's predictions were not developed rigorously, a fact which they implicitly admit by placing quotation marks around the word *derivation* when discussing their hypothesis. Furthermore, at least two of their predictions cannot be rigorously deduced from the theory, because they refer to response measures as yet unrelated to $_{S}\bar{E}_{R}$ within the theory.

While we accept the empirical relevance of many studies in which predictions have been made without logical demonstration that they flow from the theory, we hold that their theoretical relevance is problematical until such demonstration is forthcoming. The rough and ready intuitive procedures customarily employed must be replaced either by careful algebraic manipulations of inequalities following the complete chain from $_{S}H_{R}$ to $_{S}\bar{H}_{R}$ to $_{S_1+S_D}\bar{H}_{R}$ to $_{S}E_{R}$ to $_{S}\bar{E}_{R}$ to $_{S}\dot{\bar{E}}_{R}$ to the response measure, or by derivation of equations comparable to equation 17, and this step must be followed by examination of such equations in order to determine whether certain inequalities hold. It might be presumed, for example, that, in experimental conditions like those of Hovland, the theory implies that $N_1 > N_2$ should be accompanied by $\mu_{A_1} > \mu_{A_2}$; but such an inference should not be made simply by referring to Postulate 4, making the ritualistic statement that all other variables are equal, and jumping to Postulate 15. One acceptable option is to trace the effect of the inequality in N's through the chain of intervening variables listed above to μ_A. A better solution, if both N's are large enough for equation 17a to be applicable, is to determine the difference $\mu_{A_1} - \mu_{A_2}$ by successively substituting N_1 and N_2 into that equation and subtracting. If the resulting difference is in general positive for $N_1 > N_2$, then the inference is correct. Unfortunately, however, we find that the following condition must hold before $N_1 > N_2$ can be said to imply $\mu_{A_1} > \mu_{A_2}$:

$$h'\left\{mD'\left[\left(\frac{m}{50}-2\right)(10^{-iN_1}-10^{-iN_2}) - \frac{m}{100}(10^{-2iN_1}-10^{-2iN_2})\right] - \frac{c(1+k)(N_1-N_2)}{(B-W)} + \frac{c^2k(N_1{}^2-N_2{}^2)}{100(B-W)^2}\right\} > 0. \qquad [18]$$

Not only must the inference drawn intuitively be amended by imposing the restriction (equation 18), but also the drawing of this inference regarding inequalities requires additional derivations beyond that required to make the prediction of a general relationship specified by equation 17a. Thus the less specific prediction has proved the more laborious in this case.

The foregoing is intended as a warning against the misuse of the method of predicting from inequalities to inequalities. It is not, however, directed against experimental tests of predictions which are logically sound. This type of study has the special advantage that the use of data from only two experimental conditions is sufficient to permit detection of gross failures of the theory. Thus, presuming that Meehl and MacCorquodale had obtained data contrary to their predictions and that their predictions could be shown to follow from the theory, the theory would be considered unsatisfactory at that point.

On the other hand, a theory may be unsatisfactory without its delinquencies becoming apparent when tested with only two groups. For example, had Meehl and MacCorquodale deduced a theorem stating that the decrease in mean latency with increasing amounts of extramaze feeding should be linear, and had then performed a second experiment with three or more conditions, they might have found this theorem to fail to predict their new results whereas their first experiment had confirmed a less specific prediction from the same postulates. To go one step further, if the theory could be used to predict the numerical values of mean latencies as a function of amount of extramaze feeding, a third experiment designed to obtain these numerical values might fail to confirm the theory, even if the two experiments performed to test more coarse predictions had yielded confirmatory evidence.

We conclude this section by stating that, when experimental tests of predictions logically derived from Hull's theory have been reported in the psychological literature, they must be accepted as legitimate tests of the theory regardless of the level of specificity of the predictions involved. The relative merits of the three levels should not be gauged by any supposed differences in ease of prediction, but rather by the severity of the test required and by the type of experimental design and statistical analysis desired. As the degree of specificity of the prediction increases, the theory receives a more stringent test, but the restrictions upon the experimental design and the statistical analysis are likely to increase.

THE PROBLEM OF INDIVIDUAL DIFFERENCES IN PARAMETER VALUES

Earlier in this paper we assumed that all subjects given identical experimental treatment will have identical values of $_s\bar{E}_R$ and of all other parameters involved in the derivation of equation 17. Although Hull and his co-workers came close to making this assumption in their papers on the quantification of $_sH_R$ and $_sE_R$ (**3, 9**), he seems finally to have rejected it. His final views on this matter seem best to be represented by the following quotation from *Essentials of Behavior* (**7**, pp. 115–117): ". . . in the molar law of habit formation, $_sH_R = 1 - 10^{.0305N}$, . . . the value .0305 is a constant. . . . Presumably if a group of organisms strictly comparable in all other respects to those from which this equation was secured were tested on the same behavior form, a value within the sampling range of .0305 would again be secured. But if older or younger or diseased organisms were used, or if a different genetic strain or a different species, *and particularly if single organisms were used* (italics ours), presumably rather different exponential values would result."

As we understand the foregoing statement, Hull believed that the parameter values of his equations were characteristic of a population of comparable animals treated identically but not necessarily characteristic of every (or perhaps any) animal in the population. This position finds support in certain experimental work indicating that individual organisms do differ in their parameter values (**1, 2**). However, if used in this way, Hull's equations would consist of laws about populations rather than about individuals; and their potential usefulness would be greatly altered.

Two kinds of laws compared. The distinction between laws about populations and laws about individuals deserves emphasis. If a relationship is established between an experimental variable and a response variable for an individual, it is a law about an individual. If a relationship is established between an experimental variable and the mean or some other parameter of the distribution of response measures for a population of organisms, it is a law about a population.

Although laws about populations have an initial attraction to the theorist, particularly a theorist who must defend himself against the claims of the idiographic approach, they have a disadvantage which must be recognized when they are used. Because behavior takes place in individual organisms, any mechanism designed to explain behavior changes must be applicable to individuals. But since both experimental findings (**1, 2**) and mathematical reasoning (**13**) indicate that equations based upon group data may differ in form or in parameter values from the equations obtained for individuals, mechanisms inferred from population laws may prove quite unsatisfactory. When there is a discrepancy between these two types of laws, a mechanism based upon population laws may not be appropriate to the description of behavior changes in the individual. For this reason Hull's theory might lose much explanatory value if it were taken to be a system of laws about populations.

On the other hand, a law about a population has considerable actuarial value. Once the parameters of a population law have been estimated, predictions of a very practical nature may be made. Thus we could maximize the number of crossings of the Warden obstruction apparatus (**15**) by a group of sexually deprived male rats by employing a 24-hr. deprivation period, regardless of whether this period produces a maximal number of crossings in any individual animal.

Alternative solutions to the problem of individual differences. Clearly, the assumption of equal parameter values for all organisms which formed the basis for equation 16 must be rejected. If we accept Hull's solution and decide to establish laws about populations, equation 16 and subsequent equations may still be employed. We simply take, as the ${}_SO'_R$ distribution, one of inter-individual differences rather than one of intra-individual differences, and proceed as before. With this solution the theory remains testable with group data, but will suffer from the defect, previously mentioned, that it may not lead to appropriate explanatory mechanisms for the behavior of individual organisms.

The other solution is to continue to treat the ${}_SO'_R$ distribution as one of intra-individual differences and to establish laws about the behavior of individual organisms. It is to be hoped that similar organisms will exhibit similar laws, but this must be demonstrated experimentally before it may be presumed true. With this solution, data presented for a group as a whole can have little bearing upon the validity of Hull's theory. A theoretical statement of the relationship between the experimental variable and the mean response for the group will be extremely difficult because of the differences in parameter values for the different animals in the group.

A choice between these two methods of resolving our problem is not easy. Perhaps it is not necessary at this time. Research workers may use the theory either with groups or with individuals, provided that the difference between the two usages is recognized.

SUMMARY

A literal interpretation of the postulates presented in Hull's *Principles of Behavior* was shown to lead to predictions markedly different from Hull's empirical equation summarizing an experiment upon which his theory was based. Without clarification and simplification of the theory, it is believed that most predictions rigorously deduced from the theory would contradict or at least differ in form from Hull's empirical equations. A comparable statement could be made for each of the later postulate systems developed by Hull.

Three levels of specificity in prediction from Hull's theory were discussed: First, the prediction of inequalities in response measures from inequalities in experimental variables; second, the prediction of the general form of relationships between variables; and third, the prediction of numerical values for parameters in theoretical equations. The theory was seen to be most severely tested by predictions of the most specific nature. Some of the failures of Hull and others to make predictions consistent with the theory were attributed to the false belief that the deductive procedure is necessarily simplified by making predictions at a low level of specificity.

Finally, the fact that individual organisms differ in their parameter values was recognized. Two ways of theorizing which are consistent with this fact were considered: In one case Hull's equations are taken to hold for the population as a whole, in which case the theory has predictive usefulness but may lack explanatory value. In, the other case, equations are to be established for each organism separately, and will be taken to describe the process of behavior change in the individual, whether or not they lead to predictions of behavior in other individuals.

REFERENCES

1. BURKE, C. J. The function relating momentary effective reaction potential to running-time in a straight runway. Unpublished doctor's dissertation, State Univer. of Iowa, 1948.
2. COTTON, J. W. Running time as a function of amount of food deprivation. *J. exp. Psychol.*, 1953, **46**, 188–198.
3. GLADSTONE, A. I., YAMAGUCHI, H. G., HULL, C. L., & FELSINGER, J. M. Some functional relationships of reaction potential (sE_R) and related phenomena. *J. exp. Psychol.*, 1947, **37**, 510–526.
4. HOVLAND, C. I. The generalization of conditioned responses. IV. The effects of varying amounts of reinforcement upon the degree of generalization of conditioned responses. *J. exp. Psychol.*, 1937, 21, 261–276.
5. HULL, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
6. HULL, C. L. Behavior postulates and corollaries—1949. *Psychol. Rev.*, 1950, **57**, 173–180.
7. HULL, C. L. *Essentials of behavior*. New Haven: Yale Univer. Press, 1951.
8. HULL, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.
9. HULL, C. L., FELSINGER, J. M., GLADSTONE, A. I., & YAMAGUCHI, H. G. A proposed quantification of habit strength. *Psychol. Rev.*, 1947, **54**, 237–254.
10. KOCH, S. Clark L. Hull. In W. K. Estes, *et al. Modern learning theory*. New York: Appleton-Century-Crofts, 1954. Pp. 1–176.
11. MEEHL, P. E., & MACCORQUODALE, K. Drive conditioning as a factor in latent learning. *J. exp. Psychol.*, 1953, **45**, 20–24.
12. MONTGOMERY, K. C. An experimental investigation of reactive inhibition and conditioned inhibition. *J. exp. Psychol.*, 1951, **41**, 39–51.
13. SIDMAN, M. A note on functional relations obtained from group data. *Psychol. Bull.*, 1952, **49**, 263–269.
14. SIEGEL, P. S. Reactive inhibition as a function of number of response evocations. *J. exp. Psychol.*, 1950, **40**, 604–608.
15. WARDEN, C. J. *Animal motivation. Experimental studies on the albino rat*. New York: Columbia Univer. Press, 1931.

(Received October 4, 1954)

VOL. 62, NO. 5 SEPTEMBER, 1955

THE PSYCHOLOGICAL REVIEW

PRINCIPLES OF PERFORMANCE[1]

EDWARD C. TOLMAN

University of California

Apparently it behooves a so-called "cognitive theorist" such as myself to explain how his assumed intervening cognitive and motivational variables ever issue into actual behavior. I must, it seems, provide some *rules* or *principles* of *performance* according to which these variables—namely, discriminations, expectancies, drives, incentive values—add to or multiply with one another (or whatever the appropriate combining operations may be) to produce behavior. Otherwise, as Guthrie (5) has suggested, I seem to leave my animal (whether rat or human being) merely "buried in thought" (though why this should be considered so shocking in the case of a human being I have never quite been able to fathom).

Let me begin by summarizing briefly my basic assumptions. First, I assume four distinctive types of *independent variables:* (*a*) present stimulus-units; (*b*) previous presentations of stimulus-unit–response–stimulus-unit sequences, abbreviated as past (*S-R-S*)'s; (*c*) present drive disturbances and currently presented incentive objects; and (*d*) genetic or early training differences between individual subjects (as parameters affecting the results from the other three types of independent variable).

Secondly, I assume three types of *intervening variables* and *intervening-variable systems:* (*a*) discriminations within a discrimination system; (*b*) beliefs and expectancies within a belief-expectancy system; and (*c*) drives and incentive-values within a drive-incentive-value system. The three intervening variable systems are inferred sets of dispositions (in the sense of propensities of the organism to perform in certain prescribable ways under specified test conditions). And I assume that the momentary activated discriminations, beliefs, expectations, drives, and incentive values are determined by the momentary values of the presented independent variables in the contexts of these three innate and acquired dispositional systems.[2] Let me elaborate.

1. *The Discrimination System.* This is to be specified by (*a*) *categorizing* or *nominal dimensions,* that is, types of general object-classification, of which the given organism has become capable (e.g., most human beings have acquired

---

[1] Presented, with some modification, as presidential address to the Division of General Psychology of the American Psychological Association, Sept. 7, 1953.

[2] I wish to express my great indebtedness to Professor Leo Postman for his unstinting and crucial help in the development of these concepts concerning the independent and the intervening variables and variable systems. The performance principles which I shall propose later in this article, though in part Professor Postman's, are at present largely mine—at least insofar as the pictorial language is concerned. Postman does not find this language helpful.

such categories as colors, pitches, smells, chairs, tables, trees, other human beings, etc., etc.), and by (*b*) the *intensitive dimensions* along which the given organism has, in addition, become capable of making intensity discriminations (e.g., specific hues, heights of pitches, intensity of odors, comfortableness of chairs, sturdiness of tables, shadiness of trees, likeableness of human beings, etc.). Such categorizing and intensitive dimensions are uncovered in standard situations in which the organism is presented with selected arrays of stimulus-units and which it is forced (or tricked) into categorizing and/or discriminating intensitively. In other words, the discrimination system is to be discovered by the methods of a comprehensive and up-to-date, as well as a mathematically sophisticated, psychophysics. Knowing this discrimination system, we can thereafter state the specific categorizations and intensity discriminations which are likely to be activated when such and such stimulus-units are actually presented on any given occasion.

2. *The Belief-Expectancy System.* This is to be specified by the sum of the organism's expectancy dispositions. For example, a subject may show that upon the sounding of a given tone, S_1, he is prepared, provided he waits, R_1, in a given place, for the coming of a shock, S_2. Or again, he may show that when presented by the choice point of a maze, S_1, he is prepared for a resultant food, S_2, provided he turns, R_1, to the right. Or he may show that upon the presentation of a first syllable, S_1, on a memory drum he is prepared, provided he continues to look at the drum, R_2, for the coming of a second syllable, S_2. Any one such set in the animal, which (to distinguish it from the objective occurrences) may be symbolized as an $(\dot{S}_1\text{-}\dot{R}_1\text{-}\dot{S}_1)$, I would designate as an activated expectation. The continued possession by the organism of such a set I would call an expectancy disposition. An "expectancy disposition" is thus a small acquired disposition of greater or less certainty and permanence, which will tend to lead to an activated expectation whenever an instance of the given sign stimulus-unit is presented. Such an expectancy disposition will also tend to be activated by other stimulus-units which are near the sign stimulus on some common dimension of generalization. Thus, there may be groups or families of expectancy dispositions in the sense that an array of them related by different signs may all have attached to them similar expectancy dispositions as to the type of significate to be reached, if a certain general type of means-manipulatory response be carried out. Such a group or envelope of related expectancy dispositions I shall call a "belief disposition." An individual's expectancy system is constituted by the sum of all his expectancy and belief dispositions.[3]

3. *The Drive-Incentive-Value System.* This system is to be conceived as a collection of dispositions to approach and/or consume specific types of positive incentive objects or to escape from and/or avoid specific types of negative incentive objects. These approach (consummation) and avoidance (and/or escape) dispositions depend upon the interaction of two distinguishable independent variables: namely, (*a*) types and amounts of drive-disturbance, and (*b*) types of stimulus-units presented as incentive objects.

a. Drive-disturbances would be: hours of specific food-deprivation, hours of sex-deprivation in males, time-position of the female in the oestrus cycle, percentage decrement in bodily weight, hours of water-deprivation, amounts and prior time of saline injections, intensities of just-preceding electric shocks, hours of social deprivation, specific raisings and lowerings of bodily health, special maintenance-schedule with respect to aesthetic objects or with respect to exercise, etc., etc.

---

[3] I would like to emphasize that my notions about expectancies have been clarified by the recent discussion by MacCorquodale and Meehl. I wish here to acknowledge my very great indebtedness to them (13).

b. Incentive objects would be: specific foods, liquids, sex objects, groups of individuals of the same or different species, barriers, and distances to be overcome, enemies, friends, present electric shocks, poems, paintings, symphonies, open windows to look through in the case of monkeys (see Harlow [6]), etc.

The character and strength of an activated approach (consummation) and avoidance (escape) disposition will result from the combination of the particular drive-disturbance and the particular incentive object presented on the given occasion. The sum total of drive-disturbance and incentive object relations obtaining for a given individual make up his drive-incentive-value system.

Assuming now that the above brief discussion of discrimination dispositions, expectancy dispositions, and drive-incentive-value dispositions has succeeded in presenting the basic theoretical position, I wish to pass on to the main topic of this paper. How on particular occasions—given the independent variables of presented sign stimulus-units, previously presented (*S-R-S*)'s, present drive-disturbances and present positive or negative incentive objects—do the to-be-inferred resulting then-and-there activated discriminations, expectancies, drives, and incentive values interact to produce actual performance? What, in short, are the Rules or Principles of Performance which must be added to the system?

PERFORMANCES AND THE BEHAVIOR SPACE

I shall illustrate these principles or rules in large part by means of what I shall call behavior-space diagrams or pictures. This notion of a behavior space stems, of course, from Kurt Lewin's concept of the life space (12). It differs from the latter, however, *first,* in that the structures and processes which Lewin put inside his "psychological person" *as part* of the life space have been, in the present analysis, taken care of by the three dispositional systems of discriminations, belief-expectancies, and drives and incentive values, together with the momentary processes activated within these systems.

Secondly, my behavior space is more frankly a merely pictorial way of representing certain relationships and less a would-be true mathematics or logic.

Thirdly, I differ perhaps from Lewin in that what I am trying to illustrate or explain by the behavior space are not concrete responses but certain generalized *performances.* A "performance," as I here use the term, is not something that can be observed on a single occasion but something that can be specified only from observing two or more concrete behaviors. For example, Tolman, Ritchie, and Kalish in one of their experiments on spatial learning (25) used a single-unit **T** maze with two vertically opposite, alternatively usable, starting stems. And they discovered that only by observing the responses of the animals from both stems could they say whether a rat's acquired behavior—what I am here calling his "performance"—was that of "taking right-hand turns" or that of "going to the lighted side." But, whichever it was, such an "always going to the right or "always going towards the lighted side," would

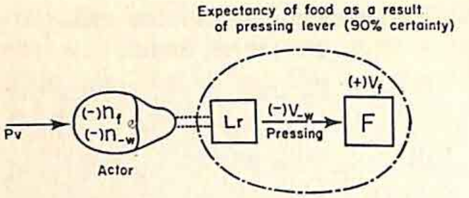

Expectancy of food as a result of pressing lever (90% certainty)
(+)V_f
(-)n_f
(-)n_-w
Pv
Actor
Lr
(-)V_-w
Pressing
F

FIG. 1. Behavior-space: rat pressing lever for food in a Skinner box.

be an example of what is here meant by a performance.[4]

Fourthly, it is to be emphasized that a behavior-space diagram will be neither a picture of the actual geographical environment nor a picture of the supposed phenomenological experience of the organism. It is rather a merely pictorial way of trying to represent the interactions between the inferred activated dispositions and their consequent determination of the to-be-predicted performance. For example, my diagram for the case of the rat pressing the lever in a Skinner box (17) at a certain rate is shown in Fig. 1.

This diagram represents the assumption that the rat discriminates himself as "actor" in the presence of the discriminated lever, *Lr*. The pair of dotted lines connecting actor to *Lr* symbolizes "immediate presence of the lever." And the diagram represents further the fact that this discriminated lever, *Lr*, has become the sign term in the activated expectancy. The whole expectancy is indicated by the contents of the dash-dot ellipse. This expectancy has acquired, it is assumed, a certainty of about 90 per cent and is to the effect that if the actor responds by the "performance" of pressing, he will get to food, *F*. Also, the expected food is depicted as having a positive valence v_f, indicated by the plus sign. And the actor is shown as having an activated drive or need-push *for* food, indicated by the minus sign, labeled n_f, inside him. Further, the length of the dashed arrow, labeled "pressing," indicates the degree of effort or work which, according to the expectancy, will be involved in the pressing. And the actor is shown as having a drive or need-push *against* work indicated by the second minus sign, labeled n_{-w} within the actor. Finally, the arrow at the left, labeled *Pv*, is a vector representing the magnitude and direction of the resulting to-be-predicted performance of pressing.

Let us consider further for a moment need-pushes and valences. I conceive a need-push as, so to speak, that portion of a drive which, under the concrete stimulus conditions of the moment, gets into the behavior space. The magnitude of a need-push is, in short, assumed to be determined not only by the magnitude of the actual activated drive but also by the goodness of the discrimination of drive stimuli under the given experimental conditions.

Consider, for example, that type of latent-learning experiment which has been called by Thistlethwaite (19) the irrelevant-incentive experiment. In such experiments, during the initial training period an irrelevant goal object (say food) is presented in one alley of a **T** or **Y** maze, and water, a relevant goal object (since the animals are in this period run thirsty), is presented in the other alley. Then the rats are made hungry and it appears from the original experiment of Spence and Lippitt (18) and from various repeat experiments (see, for example, Kendler [11]) that rats often do *not* then tend to go to the food side when first made hungry. However, it has also appeared from an experiment by Thistlethwaite (20) that, if the animals have had some previous training in discriminating their drive stimuli, they are much more likely to show some initial tendency to go at once to the food side. These two contrasting results would be interpreted, by me, as meaning that, although the rats have been made hungry (as would be shown by their readiness to approach or consume food in the standard situation), this hunger drive does not tend to be discriminated (i.e., to produce a corresponding need-push in the behavior space corresponding to the maze) unless there has been previous practice in drive discrimination in a mazelike situation. Further, some recent findings by Heyer and O'Kelly (7) indicate that tissue dehydration due to subcutaneous saline injection will not cause appropriate performance in a maze to get to water, *unless* this dehydration has been experienced over a considerable preceding period of time. I would interpret this also as lack of thirst need-push in the behavior space (except in the actual presence of water itself) unless there has been plenty of preceding time to experience and discriminate the special drive stimuli resulting from the saline-produced dehydration.

---

[4] For this notion of a performance as a behavioral disposition which can be identified only by observing at least two somewhat different test situations I am indebted primarily to Professor Benbow F. Ritchie. (Private communication.) But see also a quite recent statement of what seems to be essentially the same position by Campbell (2).

Return now to the Skinner box. We may probably assume that in it, because of preliminary training, the hunger drive stimuli and the nonwork drive stimuli have become completely discriminated. We may assume in short that the magnitude of the food need-push corresponds directly to the magnitude of the actual food hunger drive. And finally we may probably also assume that the magnitude of the nonwork need-push corresponds directly to the actual nonwork drive.

What now about the valences? Valences are conceived as bearing nearly the same relation to incentive-values that need-pushes bear to drives. That is, the amount of valence which the food pellets will have in the behavior-space diagram corresponding to the Skinner box will depend not only upon the incentive value of such pellets (as measured in some standard situation) but also upon the discriminability of the pellets in the concrete stimulus context of a Skinner box. And this discriminability may, hypothetically at least, be less than in the standard situation. Hence the valence of the food, in the behavior space corresponding to the Skinner box, could be less (though I doubt it) than the basic incentive value of such pellets, as measured in a standard situation. Similarly, the negative valence of the work involved in pressing might, in the behavior space corresponding to the Skinner box, be less than the actual negative incentive value of such work. But this also seems doubtful.

Look again at Fig. 1. My final argument is that as a result of (i) the need-push for food, (ii) the positive valence of the expected food, (iii) the need-push against work, and (iv) the negative valence of the expected work (this expected work is indicated by the length of the dashed arrow), there results the performance vector *Pv*. It is this vector which specifies the direction and magnitude of the to-be-predicted actual performance. The greater the valence of the expected food, the greater the food need-push, and the greater the expectancy that the food will result, the greater the magnitude of the performance vector toward actually pressing the lever. On the other hand, the greater the need-push against work and the greater the negative valence of the expected work and the stronger the expectation of this work, the less the performance vector toward pressing the lever.

For the sake, now, of the critic who dislikes such a pictorial mumbo jumbo, it may be pointed out that these relations probably can be equally well expressed by an algebraic equation (Equation 1).

$$Pv = f_x\ (n_f,\ v_f,\ \exp_f) - f_y\ (n_{-w},\ v_{-w},\ \exp_w) \quad [1]$$

Here *Pv* stands for the magnitude of the performance vector (i.e., the tendency to press the lever). And this magnitude is shown as increasing with the values of n_f, v_f, and $\exp_f$, combined according to some function, f_x, and as being decreased by the values of n_{-w}, v_{-w}, and $\exp_w$, combined according to some function, f_y. The exact forms of these two functions, f_x and f_y, I have not attempted to specify. It may be that these functions both have the form of simple multiplication. But in any event, it is assumed for both of these functions that the component variables all operate in a positive direction such that the increase in any one of these component variables tends to increase the value of the function.

Still further, the critic will note that the first half of the right-hand side of this equation is rather similar to Hull's equation for reaction potential, or $_SE_R$ (9), (which is presented here as Equation 2).

$$\text{(Hull)}\ {}_SE_R = f\ (D \times V_1 \times K \times {}_SH_R) \quad [2]$$

Hull's *D*, or drive, is closely similar to my n_f, or hunger need-push. His V_1, or stimulus intensity dynamism, is, however, a term I do not have. For me the effect of any *intensity* or a presented stimulus-unit such as the stimulus-pattern resulting from the lever would already have been evaluated by an analysis of the discriminatory dispositional system and of the effects in it of intense or nonintense stimulus-units upon the discriminations of *Lr* and of *F*. Hull's *K*, or incentive motivation, corresponds to my valence, v_f. And his $_SH_R$, or habit, is an *S-R* connection whereas my expectancy is conceived as an activated (S_1-R_1-S_2) expectation. Further, however, it is to be noted that Hull's equation does not contain the second part of the right-hand side of my equation. It may be recalled, however, that he does introduce, separately, the construct of conditioned inhibition $_SI_R$ which, if I interpret him correctly, is somewhat related to what I am here saying about *w* or expected work. $_SI_R$ he apparently subtracts from $_SE_R$ just as I subtract $f_y\ (n_{-w},\ v_{-w},\ \exp_w)$. It must be pointed out, however, that for Hull inhibition is somewhat different from my concept of expected work and need against work. But in any event, both he and I assume that the "expected work" or the "conditioned inhibition" gets established through repetitions of the activity involved.

Turn now to escape behavior. I would feel that, in the usual case, escape is not as has been suggested by the Hullians primarily a getting-to a positive valence attached to the reduction of pain but is rather simply a getting-away from the negative valence of the shock itself. Such an interpretation is represented by Fig. 2. In Fig. 2 the jumping is represented as a mere leading away from the grid. And the animal will merely jump in any direction easily available.

Turn next to avoidance behavior. Avoidance behavior has always been something of a problem for *S-R* psychologists. This point was made some years ago by Hilgard and Marquis (8) and has recently been re-emphasized and subjected to a further analysis by

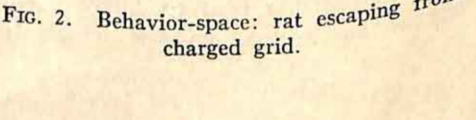

Expectancy of continued shock from being on grid (90% certainty)
(-)V_sh
Shock
(-)V_-w
Being on grid
Grid
(-)n_-sh
(-)n_-w
Pv
Jumping off

FIG. 2. Behavior-space: rat escaping from charged grid.

Ritchie (16). In avoidance behavior the performance vector results, according to my language, from the expectation that going to the grid is going to lead to shock. And this performance vector pushes or pulls the animal away from the grid even when he is not on it. The negative need-push against shock will lead, not as in the escape case merely to getting off the grid after the shock has come, but rather to the performance of avoiding the grid altogether. See Fig. 3. In such a case, pain never occurs and hence there is no pain to be reduced. Mowrer (14), of course, is the one among the stimulus-response-oriented psychologists who has most clearly seen this. And he has solved it by postulating a conditioned anxiety to the whole grid box, the reduction of which constitutes positive reinforcement. Now while I would accept something like Mowrer's interpretation for many instances, I do not believe that anxiety is always or necessarily involved. The skilled human subject who has learned to lift his finger in time to avoid a minor shock is not usually, I believe, suffused with even a minimum of anxiety. He responds so as to avoid the shock, quite unemotionally and purely on the basis of his expectancy, his need-push, and the negative valence of the shock. He expects shock in a certain place at a certain time, provided he goes to or stays in that place, and his performance is one of avoiding that place before that time, in a quite relaxed, nonanxious fashion.

Finally, however, it is to be noted that although my *diagrams* for escape and avoidance are different from that for approach, my algebraic equation for either escape or avoidance would be of the same form as that for approach (Equation 3).

$$Pv = f_x\,(n_{-sh},\ w_{-sh},\ \mathrm{exp}_{sh}) - f_y\,(n_{-w},\ v_{-w},\ \mathrm{exp}_{w}) \qquad [3]$$

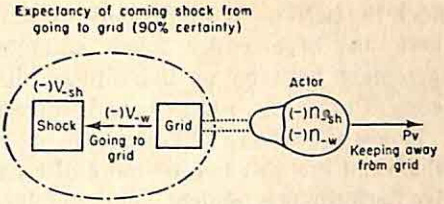


FIG. 3. Behavior-space: rat avoiding grid from which shock is expected.

The greater the need-push n_{-sh} against shock, the greater the negative valence v_{-sh} of the particular shock and the more certain the expectation exp_{sh} (or immediate discrimination of shock), the stronger will be the performance vector leading to behavior away from the sign of pain (in the case of avoidance) or leading to the leaving of the pain-object (in the case of escape). And the greater the need against work n_{-w}, the greater the negative valence v_{-w} of the particular work involved in the escape or the avoidance and the greater the expectancy of this work, the smaller will be the *Pv* toward escape or avoidance.

It is to be noted further, however, that, in the avoidance case, *w*, or the expected work involved between sign and significate, can be large and lead to relatively weak avoidance of the sign as a result of a number of relatively different physical reasons. The *w* may be waiting a long time for the shock. Or it may involve doing some special further act such as pressing a treadle for the shock to appear. Or, to take a quite different example, where the significate is not pain from shock but pain from a rat-hungry cat, *w* might involve the breaking down of a tough barrier by the rat between it and the cat. In all three cases the avoidance response of the rat will be relatively weaker than it would have been if these intervening *w*'s had been small.

But up to now you may well feel that my formula and Hull's are not significantly different. And, if you do, so much the better. It would indicate that there has been really much common agreement between us in spite of differences of words, phrases, and slogans.

However, when it comes to cases where not one but two or more alternative performance tendencies are involved and result in some quite new performance—such, for example, as taking a compromise path between two original ones, taking a shortcut, taking a roundabout route, VTEing, or the like—then depicting the situation by means of the behavior space becomes (to me) more helpful and suggestive. (And, of course, not to renege on my basic position, my own "private style of thinking," to quote Kendler [11], is such that I cannot easily think about the interaction of a single sign, expectancy, need-push, and valence without first seeing these in some sort of spatial terms.) But turn now to the new problem of combining two expectancies. I shall introduce here a new concept, that of the angular dispersion between expectancies.

This concept of angles is tricky and I know I am not wholly clear on it. It is one that Lewin also bogged down on. Nevertheless let me present, as an illustration, one type of compound performance which I would explain with the help of angles, namely, VTEing (vicarious trial and error). The explanation will be closely related to, but in some ways different from, that which Clark Crannell and I evolved some years ago (21, 22) in connection with what I then called the "schematic sowbug." Actually, anyone who read the earlier papers may have already recognized that the actor as I have presented him here is no more than a modern, and I hope a less complicated, version of my old friend, the schematic sowbug.

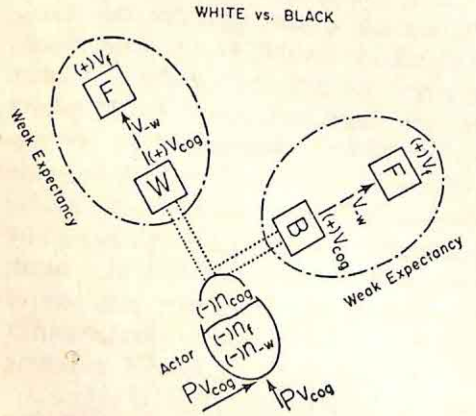


FIG. 4a. Behavior-space, early in discrimination learning. White card versus black card. Much VTE.

There are two sets of empirical findings as to VTEing to be emphasized: (*a*) Early in learning, just before (or just as) the rats begin to react to the crucial sign dimension, they tend to do more VTEing when there is a large discriminable difference between the two cards than when there is a small one. They VTE more for a white card vs. a black card than for a white card vs. a light gray card (21). (*b*) Later, however, when they have had much preceding experience in discriminating along the white-gray-black dimension, they tend to VTE more when the discriminable difference between the two cards is small (24). They VTE more for white vs. light gray cards than for white vs. black cards.

Consider first the behavior-space pictures early in learning when the rats VTE more for the white-black than for the white-gray. (See Fig. 4a and Fig. 4b.)

In Fig. 4a the actor (the sowbug) is in the white-black situation and in Fig. 4b in the white-gray one. And I have assumed that white vs. black corresponds to a 90° angle relative to the actor's (sowbug's) nose, whereas the white vs. gray corresponds to a smaller angle. I have also assumed that in this early part of learning there operates, in addition to the hunger need-push n_f, a more purely cognitive sort of need-push to examine new stimuli, n_{cog}. The need-push corresponding to this cognitive-examination drive I have indicated by a minus sign in the fore part of the actor or bug. I have also assumed that both the stimulus cards have for a while, as new stimuli, positive valences for this cognitive-examining drive. That is, there will be performance vectors pushing the actor to "look at" each stimulus card. These vectors are those indicated at the rear of the bug. In addition, I have indicated, by putting in F (for food) with a plus valence on it on both sides, that at this early stage of learning there have developed expectancies that going toward either card may lead to food. And for the diagram to be complete there should also be another pair of shortish vectors pushing the bug, as a result of these expectancies, to jump towards each card. I have omitted these latter, jumping vectors, so as not to complicate the picture. The vectors shown are "looking towards and examining" vectors.

I shall assume further (and here I am assuming very much what Glanzer [3, 4] has proposed in two recent articles in which he was concerned with alternation behavior in a symmetrical maze) that in the course of looking at one card the valence of this card as "something to look at and examine" tends to become temporarily discharged or satiated. This then weakens, for the moment, the vector for looking at this first card. The actor then turns to look at the other card. The vector toward this latter card, however, then becomes satiated in its turn; whereas the vector for looking at the first card becomes stronger again. So the rat looks back and forth, or VTE's, until satiation for both cards, as far as looking or examining is concerned, has become complete. In short, the rat VTE's until jumping takes over.

To explain further why there is more VTEing for white vs. black (Fig. 4a) than for white vs. gray (Fig. 4b) I shall also assume that since in the case of white vs. black the two cognitive-examination vectors are at right angles, the reduction of one of these vectors through satiation has no effect on the other. The rat gets temporarily satiated for white but he still wants to look at black, and vice versa. In the case of the white vs. gray, on the other hand, the two examining vectors are at a narrower, acute, angle which means that the shortening of one vector through satiation projects upon and tends to shorten the other. Looking at white reduces temporarily not only the vector to look at white but also the one to look at gray, and vice versa. Hence early in learning or just before learning there will be less VTEing for white vs. gray than for white vs. black.

Consider now the situation later in learning. The original cognitive-examination drive is assumed now to have become wholly satiated. The vectors corresponding to it will have disappeared. But expectations have begun to build

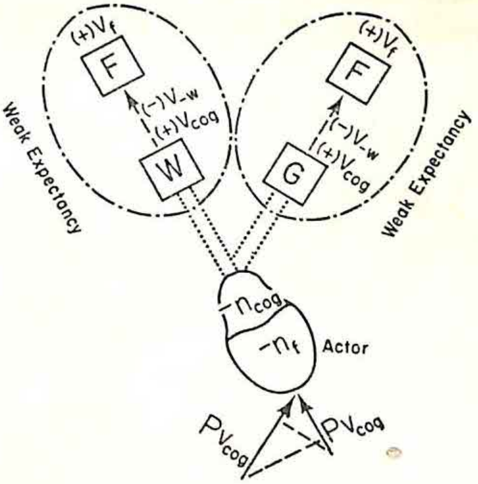


FIG. 4b. Behavior-space, early in discrimination learning. White card versus gray card. Little VTE.

up to the effect that, in the one case, white leads to food and black to no food, and, in the other, that white leads to food and gray to no food. These situations are represented in Figs. 5a and 5b.

The vectors are no longer ones for looking, but ones for jumping to white (as sign of food) and for not-jumping to black (as sign of nonfood). In the white-black case (Fig. 5a), because the two vectors are at right angles the negative vector "away from the black" does not affect the positive one "toward the white," and vice versa. In the white-gray case (Fig. 5b), on the contrary, since the two arrows are at an acute angle, the negative vector "away from the gray" does tend to project upon and reduce the positive vector "toward the white"; and the positive one "toward the white" projects upon and decreases the negative one "away from the gray." And this means in the white-gray case that until the expectancies become very certain and the vectors very long, the reduction of the positive vector by the negative one and vice versa will be sufficient to keep the rat in unstable equilibrium so that VTEing will tend to continue much longer.

Now, for the sake of the nonvisually minded critic, let me here suggest that perhaps a more respectable and conventional way of talking would be to say merely that early in learning there is more generalization of satiation of the cognitive examining drive from white to gray than from white to black; whereas, later in learning, there is more generalization of acquired approach and avoidance tendencies from white to gray than from white to black. More generalization of satiation for "looking at" early in learning means *less* VTEing between white and gray, whereas more generalization of approach and avoidance tendencies later in learning means *more* VTEing between white and gray. Actually, however, I suspect that such a more "respectable" way of talking is really just as confused and as full of hidden assumptions as is my pictorial way. In any event, the pictorial terms suggest to me (and I emphasize *to me*) a greater number of further experiments to try out than do the purely verbal ones.

The topic of this paper was "Principles of Performance." And by a per-

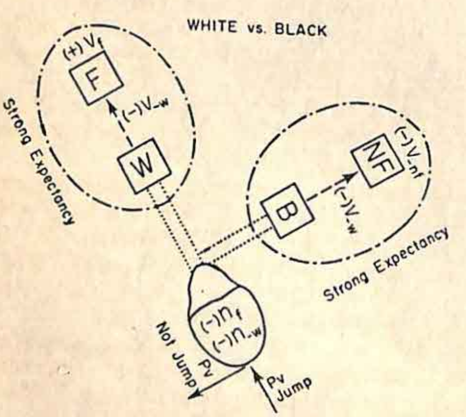


FIG. 5a. Behavior-space, later in discrimination learning. White card versus black card. Little VTE.

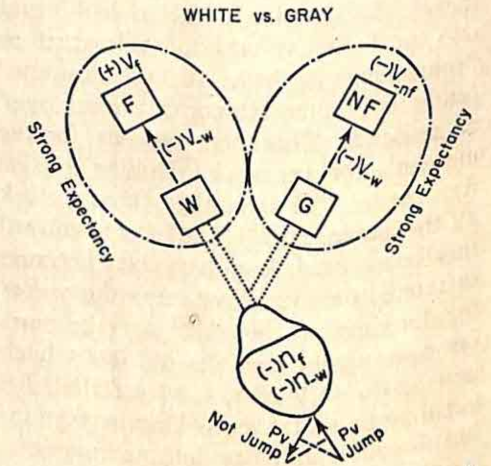


FIG. 5b. Behavior-space, later in discrimination learning. White card versus gray card. Much VTE.

formance was meant, as has been said, a generalized way of behaving to be discovered and specified only by observing the responses in at least two and perhaps more concrete test situations. Further, a performance when thus discovered is identifiable only in general terms, such as approach, consummation, escape, avoidance, vicarious trial-and-error examination, or the like. The motor skill characteristics involved in such a performance are not included in its definition. Only in a very general way may a performance be further specified as a "looking at," a "jumping to," a "running down," a "pressing," an "eating," or a "drinking." The point of view being argued for is that performances and not movements or skills are all that we ought now, at the present time, to try to explain or to predict in detail. Under certain controlled conditions we can use movements as pointer readings for performances but we cannot as yet predict movements independently of the performances within which they occur. As Brunswik has pointed out (1), all kinds of different movements and skills may vicariously substitute for one another and produce one and the same achievement or performance. Movements and skills in his terms are mere mediational, peripheral phenomena. It is the "distal," not the "proximal," characters of *S*'s and *R*'s to which I am suggesting that we devote ourselves for the time being. This is my interest and my bias.

One last word. Most of the really interesting problems of psychology—such as the distortion effects of motivation on perception (and I still believe there are some) or again those facts which have to do with what I have called the widening or narrowing of cognitive maps (23)—will eventually find their place, I believe, under the heading of complications in the principles of performance. Or, to use the language of this paper, they will turn out to be describable in terms of interesting complications and interactions between the performance vectors, as these complications and interactions appear in particular behavior-space setups.

REFERENCES

1. BRUNSWIK, E. Organismic achievement and environmental probability. *Psychol. Rev.*, 1943, **50**, 255–272.
2. CAMPBELL, D. T. Operational delineation of "what is learned" via the transposition experiment. *Psychol. Rev.*, 1954, **61**, 167–174.
3. GLANZER, M. The role of stimulus satiation in spontaneous alternation. *J. exp. Psychol.*, 1953, **45**, 387–393.
4. GLANZER, M. Stimulus satiation: an explanation of spontaneous alternation and related phenomena. *Psychol. Rev.*, 1953, **60**, 257–268.
5. GUTHRIE, E. R. *The psychology of learning* (Rev. Ed.). New York: Harper, 1952.
6. HARLOW, H. F. Motivation as a factor in the acquisition of new responses. In *Current theory and research in motivation, a symposium.* Lincoln: Univer. of Nebraska Press, 1953, pp. 24–49.
7. HEYER, A. W., JR., & O'KELLY, L. I. Studies in motivation and retention. *Comp. Psychol. Monogr.*, 1951, **20**, Serial No. 106.
8. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: Appleton-Century-Crofts, 1940.
9. HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
10. KENDLER, H. H. An investigation of latent learning in a T-maze. *J. comp. physiol. Psychol.*, 1947, **40**, 265–270.
11. KENDLER, H. H. "What is learned?" A theoretical blind alley. *Psychol. Rev.*, 1952, **59**, 269–277.
12. LEWIN, K. *A dynamic theory of personality.* New York: McGraw-Hill, 1935.
13. MACCORQUODALE, K., & MEEHL, P. E. Section 2 in Estes, W. K. *et al. Modern learning theory.* New York: Appleton-Century-Crofts, 1954.
14. MOWRER, O. H. On the dual nature of learning—a reinterpretation of "conditioning" and "problem-solving." *Harvard Educ. Rev.*, 1947, **17**, 102–148.
15. MURRAY, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.

16. RITCHIE, B. F. Can reinforcement theory account for avoidance? *Psychol. Rev.*, 1951, **58**, 382–386.

17. SKINNER, B. F. *The behavior of organisms: an experimental analysis.* New York: Appleton-Century-Crofts, 1938.

18. SPENCE, K. W., & LIPPITT, R. An experimental test of the sign-gestalt theory of trial-and-error learning. *J. exp. Psychol.*, 1946, **36**, 491–502.

19. THISTLETHWAITE, D. A critical review of latent learning and related experiments. *Psychol. Bull.*, 1951, **48**, 97–129.

20. THISTLETHWAITE, D. Conditions of irrelevant-incentive learning. *J. comp. physiol. Psychol.*, 1952, **45**, 517–525.

21. TOLMAN, E. C. Prediction of vicarious trial and error by means of the schematic sowbug. *Psychol. Rev.*, 1939, **46**, 318–336.

22. TOLMAN, E. C. Discrimination vs. learning and the schematic sowbug. *Psychol. Rev.*, 1941, **48**, 367–382.

23. TOLMAN, E. C. Cognitive maps in rats and men. *Psychol. Rev.*, 1948, **55**, 189–208.

24. TOLMAN, E. C., & MINIUM, E. VTE in rats: overlearning and difficulty of discrimination. *J. comp. Psychol.*, 1943, **36**, 91–98.

25. TOLMAN, E. C., RITCHIE, B. F., & KALISH, D. Studies in spatial learning: II. Place learning versus response learning. *J. exp. Psychol.*, 1946, **36**, 221–229.

(Received November 22, 1954)

Psychological Review
Vol. 62, No. 5, 1955

A NOTE ON SECOND-ORDER CONDITIONING—AND SECONDARY REINFORCEMENT

GREGORY RAZRAN

Queens College

In *Principles of Behavior* (12, ch. 7), Hull set up second-order conditioning as a sort of fulcrum of his much used construct of secondary reinforcement. And, except for Skinner (30, p. 245), other CR learning theorists seem to regard it, too, as quite a general CR manifestation (19, 22, 28, 31). Indeed, one gets the impression that most of these theorists subsume second-order conditioning under secondary reinforcement in the sense that an S-R or R that is able to reinforce another S-R or R must *ipso facto* be able to condition (in a classical manner) the other S-R or R to itself (19, pp. 438, 457; 22, pp. 430–433; 28; 31, pp. 697–699). Frolov's experiment (7), reporting that the sight of a black square paired with a conditioned sound of a metronome came to produce by itself conditioned salivation, is widely quoted in textbooks and in integrative summaries and reviews of the topic.

Yet the fact remains that, as directly observed in the laboratory, second-order conditioning is a rather uncommon and unstable phenomenon. Of the 347 Russian reports on classical conditioning of dogs and monkeys that the writer has managed to see only eight (4, 6, 7, 9, 14, 16, 17, 33) [1] offer accounts of any degree of successful second-order conditioning. Much more commonly—in 63 separate investigations—nonreinforced pairings of conditioned and nonconditioned stimuli are reported as resulting in conditioned inhibition; that is, the second-order to-be-conditioned stimuli, instead of evoking a CR come to interfere with it when they are applied simultaneously or shortly before the first-order conditioned stimuli. (In an even larger number of cases, as might be expected, the Russian protocols show that the pairing produces no effect at all; or, as Pavlovians usually put it, "the animal's nervous system was too excitable" [or insufficiently inhibitable —one would seem to have it either way in that system] "to develop the conditioned inhibition.") True, a greater frequency of second-order conditioning may be found in Russian experiments with lower animals (hermit crabs 18; fish 8); with school children (10, 20, 23, 29—more often with Ivanov-Smolensky's instrumental than with Krasnogorsky's classical technique); and with adult human subjects (1, 13, 15, 21). But even here, results are at times sporadic and the finding itself poses some special questions. Moreover, no bona fide second-order conditioning has ever come out of any American laboratory,[2] and the writer has never been able to demonstrate it in any of his series of studies with "misinformed" subjects in whom most other behavioral Pavlovian findings have been corroborated.

---

[1] Reference to the original reports of the experiments by Froloff and by Lindberg, cited respectively in Pavlov's *Conditioned Reflexes* (7, p. 27) and in Pavlov's *Pavlovian Wednesdays* (16, pp. 196, 201, 224, 241), could not be found in any Russian bibliographical periodical. The writer assumes that such reports have never been published.

[2] The experiments by Culler and Finch (5) and by Brogden and Culler (2) are a somewhat special case, inasmuch as the higher-order to-be-conditioned stimuli were combined with a thoracic shock as a "motivator" or reinforcer (pedal shock was the unconditioned stimulus).

If second-order conditioning is thus not a general but a very special CR manifestation, two hypotheses would seem to be suggested by the studies in which it reportedly has manifested itself. One—which the writer intends to reject—regards second-order conditioning as merely an artifact of reactivated CR *stimulus generalization* or, as Woodworth and Schlosberg term it, CR *stimulus nondifferentiation*—a more descriptive term than the writer's CR stimulus *pseudo-generalization*. It might, for instance, be argued that, while nonreinforced applications of nonconditioned stimuli by themselves improve, through contrasts, CR stimulus differentiation, nonreinforced combined applications of nonconditioned and conditioned stimuli (method of second-order conditioning) on the contrary reactivate, through confusion, CR stimulus nondifferentiation. Hence: second-order stimuli may well evoke CR's for which they originally were ineffective. Reactivated CR stimulus nondifferentiation is quite frequently reported by Russian laboratories when subjects are administered some drug or undergo surgery or succumb to illness or are merely confronted with a difficult task of differentiation (also, after a lapse of time)—and presumed second-order conditioning might conceivably be another example of it.

To be sure, such an assertion could be readily checked by series of CR nondifferentiation tests before and after second-order training sessions. But, unfortunately, no account of such tests could be found in the Russian CR reports, and their finding of more second-order conditioning in lower animals (whose capacity for differentiation and for learning in general is admittedly underdeveloped) than in dogs and monkeys would seem to indicate further an "artifact-nondifferentiation" view. Yet this view is strongly opposed by two separate sources of evidence. (*a*) In three otherwise well-controlled Russian studies—two with dogs (4, 16) and one with monkeys (17)—second-order conditioning persisted in its original form even after its first-order conditioning was "transswitched," i.e., changed from food to defense CR's, and vice versa. (*b*) All second-order CR's in dogs and monkeys were best formed—whenever they were formed at all—when their first-order CR's were well established and hence at a minimum of stimulus nondifferentiation. Consequently, the hypothesis of second-order conditioning as an artifact and a case of stimulus nondifferentiation might better be rejected as an over-all interpretation of the phenomenon, even though it might still be retained as a plausible account of its alleged manifestation in fish and invertebrates.

The second hypothesis, taking its cue from Russian CR studies of school children, would maintain that second-order conditioning, unlike first-order conditioning, requires some sort of organismic cognition to be established. And here the writer would like to present direct evidence from one of his own experiments. In this experiment, food was used as the unconditioned stimulus, the sound of a metronome and the flashing of a green light were the first-order and second-order to-be conditioned stimuli, and salivation was the unconditioned reaction. The 15 adult human subjects in the experiment were divided equally into a "misinformed" group and two "informed" ones: one "informed without suggestion" and the other "informed with suggestion." The misinformed group was told in the first-order conditioning that "the purpose of the experiment is to study the effects of auditory fatigue upon digestion" and in the second-order conditioning that they were "tested to find out which affects digestion more, auditory or visual stimuli." On the other hand, the informed-without-suggestion group was instructed about Pavlov's work with dogs and told that "the purpose of the experiment is to find out whether human subjects condition similarly," while the informed-with-suggestion group was told that human subjects *do* condition like Pavlov's dogs and that "the purpose of the experiment is to gather more data on the subject."

The results—scores in terms of *t* values between conditioned and control salivation—are presented in Table 1 and are quite clear-cut. *The informed groups, particularly the informed-with-suggestion group, manifested a considerable amount of second-order conditioning, whereas the misinformed group, whose first-order conditioning was by all tokens more regular and greater than that of the informed groups did not.*

TABLE 1

Second-Order Salivary CR's of 15 Adult Human Subjects as Dependent upon Their Cognition of the Conditioning

(Scores are *t* values (*df* = 35) between conditioned and control salivations, and are in all cases results of three 50-min. training sessions.)

Stimuli Tested	Groups Tested		
	Misinformed	Informed Without Suggestion	Informed With Suggestion
1. Metronome (first-order stimulus)	24.5	2.9	15.4
2. Green light (second-order stimulus)	.6	1.9	3.6

The hypothesis that second-order conditioning needs cognition to be adequately effected is not for the present concerned with the nature of the cognition needed. The distinction between cognition and noncognition is so far a purely global-operational one: the differences obtained between subjects who know and who do not know that they are being conditioned. Specific conceptualizations of processes involved in the cognition might better be held in abeyance until after specific experiments have been devised to test them. Moreover, the hypothesis does not of course deny such cognition to animals. All it maintains is that in typical experiments of classical conditioning, animal subjects manage only occasionally to exercise adequate cognition while as a rule they fall short of it simply because they do not cognize what their human experimenters want them to cognize—or do. (In instrumental or operant conditioning, in the writer's estimation, animal subjects "catch on" more readily since there, as often stated, the to-be-conditioned or to-be-reinforced S-R or R produces, and not merely accompanies, the unconditioned or reinforcing S-R or R.)

Elsewhere (25) the writer has postulated—upon the basis of an array of experimental evidence—a central cognitive factor in stimulus generalization; and here, in second-order conditioning, such a factor seems to him even more likely to be operative. And in both cases, it needs to be repeated, the argument rests upon the same two convergent considerations: (*a*) the differences between "knowing" and "unknowing" human subjects (partly also the "introspections" [sic!] of the subjects) with respect to the CR characteristic in question as found by the writer; and (*b*) the irregularities in the characteristic as evident in the large number of studies with animal subjects in the Russian laboratories. Basic irregularities in a phenomenon like conditioning, the data of which come from several hundred well-controlled experiments of a uniform design, cannot, the writer posits, be resolved by merely negating the phenomenon or ignoring the irregularities. Rather, they call for supplementary assumptions, which is what is proposed here.

Now, if second-order conditioning is difficult to obtain and maintain in animals, and it seems to need cognition to be set up in human beings, what about secondary reinforcement per se, conceived independently as the capacity of a conditioned stimulus to reinforce more or less permanently another reaction? In the writer's opinion, the answer is essentially the same. True, the laboratory evidence for secondary reinforcement in animals—while never too critically appraised—seems, to date, to be more consistent than that for second-order conditioning. Yet, as an isolated force, secondary reinforcement is, too, as any mere inspection of the data of its several experiments[3] reveals, very impermanent, very extinguishable, and quite ancillary and immediately dependent upon primary reinforcement. Thus there is here, too, a far cry from the results obtained (several experiments) to the phenomena (the vast area of man's cultural behavior) which are purported to be explained—and lots of room is left for supplementary accounts.

Twenty years ago, in a critical analysis of the entire CR area in phylogeny and ontogeny, the writer suggested what he then called a "reservoir theory" of relative CR permanence and "unextinction"—viz.: ". . . the emergence somewhere in the evolutionary scale of an organismic pattern, perhaps the cortical representation of the vocal or proprioceptive parts of the conditioning [unconditioned] and conditioned stimuli, that acts as a reservoir of associational 'energy' and makes the strength and existence of a connection or C-R independent of the 'energy of' or 'reinforcement by' the original stimulus that caused the C-R or connection to be formed" (24, p. 108). And more recently, in studying the "interiorization of conditioned attitudes," he found that "to attain some permanence, to stay conditioned, conditioned attitudes require some post-conditioning verbalization which seems to act as some sort of consolidator of the conditioned change" (26, p. 3). Consequently, it would seem reasonable to assume that secondary reinforcement, too, may fall within the category that needs verbalization and/or cognition to become a part of man's permanent behavior repertoire.[4]

On the other hand, the writer would contend that conditioning in which the unconditioned reaction or the primary reinforcer is a dominant biological activity—that is, primary or first-order classical conditioning or primary instrumental or operant conditioning—need not invoke cognition for its initiation and maintenance (even if cognition may well affect its operation). The basis for such a contention is plain. Classical conditioning is quite effective in a variety of instances in which the assumption of cognition would be patently gratuitous. Consider, for instance, the conditioning of decorticated and curarized animals (11), of paramaecia (32), of coelenterates (34), and, above all, the large volume of successful conditioning of all types of visceral reflexes, including visceroviscral ones, that the Russians have reported in recent years (cf. 3, 27). And, while there is no evidence for "primitive" primary conditioning of the instrumental-operant type, there would seem to be—at least for the present—no methodological advantage and no "ontological argument" for invoking cognition as the *sine qua non* for its coming into being.

---

[3] The "reinforcement" of secondary reinforcement must obviously go beyond the mere occurrence of the conditioned reaction, thus excluding a number of experiments that have been bracketed with it. A guinea pig running away from a cage in which he had been shocked, or a dog retrieving an object that had been associated with food, or even a college student reporting a food sensation to a conditioned metronome, are cases not of secondary reinforcement but of primary conditioning

[4] It need also be mentioned that nonclassical reinforcement learning (or conditioning) admittedly takes place in settings in which, compared with classical conditioning, cognition is more likely to be available and more likely to be utilized (skeletal vs. autonomic action; choice, or at least more choice of, response).

But all this does not in any way negate the hypothesis of the cognitive nature of second-order classical conditioning and of secondary reinforcement. On the contrary, it corroborates a view that the writer has held for some time—namely, that the learning of anything, whether it be a movement or a motive or an act or an attitude, must depend directly on either some source of specific biological energy or some center of general cognitive direction (whatever the ultimate nature of the energy—or the correlated energy—of the latter). In learning, the writer is an unabashed dualist.

REFERENCES

1. BEKHTEREV, V. M., & DOBROTVORSKAYA, N. I. [The development of chained associated reflexes.] *Vopr. Izuch. Vospit. Lichn.*, 1927, 3–4, 1–10.
2. BROGDEN, W. J., & CULLER, E. Experimental extinction of higher order reflexes. *Amer. J. Psychol.*, 1935, **47**, 663–669.
3. BYKOV, K. M. [*The cerebral cortex and the internal organs.*] Moscow: Medgiz, 1947.
4. CHEBOTAREVA, O. [Further materials in the physiology of conditioned reflexes.] Thesis, St. Petersburg, 1912.
5. CULLER, E., & FINCH, G. Higher order conditioning with constant motivation. *Amer. J. Psychol.*, 1934, **46**, 596–602.
6. DOROZDENKO, N. P. [The effector path of conditioned reflexes of the second order.] *Fiziol. Sh. SSSR.*, 1950, **36**, 519–523.
7. FROLOV, I. P. Quoted in I. P. Pavlov, *Conditioned reflexes*. New York: Oxford Univer. Press, 1927. Pp. 34–35.
8. FROLLOF, J. P. [Frolov, I. P.] Bedingte reflexe bei Fischen, II. Mitteilung. *Pflüg. Arch. Physiol.*, 1928, **220**, 339–349.
9. FURSIKOV, D. S. [The relation between the processes of excitation and inhibition.] *Trud. Fiziol. Lab. Pavlova*, 1924, **1**, 1–47.
10. GACKEL, L. [An investigation of conditioned inhibition in hysterical children.] *Zh. Nevropatol. Psikhiat.*, 1928, **21**, 682–690.
11. GIRDEN, E. Conditioned responses in curarized monkeys. *Amer. J. Psychol.*, 1947, **60**, 571–587.
12. HULL, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
13. ISRAELSON, Z. I. [Topical differentiation of associated motor reflexes to dermal and muscular stimuli.] *Izv. Imper. Voyenno-Med. Akad.*, 1914, **28**, 73–78.
14. KASHERININOVA, N. A. [Contributions to the study of salivary conditioned reflexes to tactile stimuli in the dog.] Thesis, St. Petersburg, 1908.
15. LENZ, A. [Conditioned reflexes of higher order and their study in mental patients.] *Russ. Fiziol. Zh.*, 1922, **5**, 93–.
16. LINDBERG, A. A. Quoted in I. P. Pavlov, [*Pavlovian Wednesdays.*] Moscow and Leningrad: Akademiya Nauk, 1949. Vol. 1: pp. 196, 201, 224, 240; Vol. 2: p. 229.
17. MALINOVSKY, O. V. [Conditioned reflexes of the second order in monkeys.] *Trud. Inst. Fiziol. Pavlova*, 1952, **1**, 205–212.
18. MIKHAILOFF, S. Expériences réflexologiques. Expériences nouvelles sur Pagurus striatus, Leander xiphias et treilleans. *Bull. Inst. Oceangr.*, 1923, No. 422, 1–16.
19. MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 435–472.
20. NOVIKOVA, A. A. [Conditioned inhibition and conditioned reflexes of "higher order" in children.] *Med.-biol. Zh.*, 1929, **1**, 120–131.
21. OPARINA, N. [The relations of the processes of excitation and inhibition during the action of extra stimuli.] *Nov. Refleksol. Fiziol. Nervn. Sist.*, 1929, **3**, 196–209.
22. OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford Univer. Press, 1953.
23. PANFEROV. Quoted in N. I. Krasnogorsky, Bedingte und unbedingte Reflexe in Kindesalter und ihre Bedeutung für die

Klinik. *Ergebn. inner. Med. Kinderhk.*, 1931, **39**, 613–730.

24. RAZRAN, G. Conditioned responses: an experimental study and a theoretical analysis. *Arch. Psychol.*, 1935, No. 191.
25. RAZRAN, G. Stimulus generalization of conditioned responses. *Psychol. Bull.*, 1949, **46**, 337–365.
26. RAZRAN, G. The interiorization of conditioned attitudes. Paper read at East. Psychol. Ass., New York, 1954.
27. RAZRAN, G. Conditioning and perception. *Psychol. Rev.*, 1955, **62**, 83–95.
28. SEWARD, J. P. Secondary reinforcement as tertiary motivation: a revision of Hull's revision. *Psychol. Rev.*, 1950, **57**, 362–374.
29. SHASTIN, N. R. [Unconditioned and conditioned reflexes in myxedema.] *Russ. Fiziol. Zh.*, 1930, **6**, 470–482.
30. SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.
31. SPENCE, K. W. Theoretical interpretations of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 690–729.
32. TCHAKHOTINE, S. Réactions conditionnés par micro-puncture ultra-violette dans le comportement d'une cellule isolée (*Paramecium caudatum*). *Arch. Inst. Prophyl.*, 1938, **10**, 119–133.
33. ZELIONY, G. P. Espèce particulière de réflexe conditionnel. *Arch. Sci. Biol., St. Petersburg,* 1909, **15**, 437–453.
34. ZUBKOV, A. A., & POLIKARPOV, G. G. [Conditioned reflexes in coelenterates.] *Usp. Sovr. Biol.*, 1951, **32**, 301–302.

(Received October 8, 1954)

Psychological Review
Vol. 62, No. 5, 1955

PSYCHOLOGY FROM A BIOLOGIST'S POINT OF VIEW

C. JUDSON HERRICK

Grand Rapids, Michigan

THE BASIC PROBLEM

A student of the philosophy of science might be tempted to say there are as many biologies as there are biologists, if account is taken only of the men whose exceptional insight and productiveness have guided the growth of the science. The same may be true of psychology, and perhaps we can do no better than accept Cattell's characteristic definition—"Psychology is what the psychologist is interested in *qua* psychologist." There are, however, some general principles about which biologists and psychologists are in substantial agreement. One of these principles is that no factual findings have scientific significance until they are fitted into the appropriate niche in the integrated system of knowledge.

The neurologist finds this orientation especially difficult because almost all experiences and activities of men and other animals involve nervous functions and his field has no boundaries. The human brain is the most important thing in the world, for, as Gibbs (1) expresses it, "Human history is a history of the brain activity of the human race" (p. 1505). This relationship ties neurology closely with psychology and also with psychiatry, sociology, and every other human interest. But when the neurologist tries to find out just where his findings tie in with psychology he is puzzled. There are so many psychologies that one wonders what it is all about.

The interested spectator who sits on the fence watching the game sees two opposing teams and, on the side lines, a goodly number of other psychologists who do not join either faction. In one team the partisans of traditional dualism contend for a sharp separation of the conscious, or "spiritual," activities from the unconscious, or "physical," thus splitting the world as we experience it into two universes, one of which has been characterized as "spiritual reality" or "ideational reality" and the other as "physical reality." Opposed to these radical spiritists are the mechanists, who insist that, since the search in both science and philosophy is for unifying principles, and since it has not been possible to explain how a nonphysical agency can act upon a physical structure so as to influence human conduct, we must search for physicalistic principles of sufficiently wide import to embrace all the known phenomena.

The more radical members of the second group, apparently accepting the traditional doctrine that anything "spiritual" is *ipso facto* nonphysical, take the easy way out and deny that conscious experience of any kind has scientific or operational significance. This despite the fact that the very denial is a conscious act. This exclusion of everything mentalistic from psychology is obviously a defense reaction against the primitive animistic mythologies which still survive in every human culture. But even though mind is called an epiphenomenon, it is nonetheless a phenomenon, a natural event, and a place in the system of nature must be found for every natural event.

The spiritists' quest for a psychology released from the limitations imposed by the laws of the physical world, and the objective psychologists' insistence that only observable physical processes are significant for psychology, seem to be irreconcilable. The controversy points again to the fact that the basic problem of psychology is, as it always has been, the exact nature of the relationship between our knowledge of the objective world and the subjective experience of knowing and all other conscious acts.

My purpose here is to examine, from the standpoint of mechanistic biology, some of the diverse fields of inquiry which must be integrated before this basic problem can be formulated in accordance with physical, physiological, and psychological principles now generally accepted. The divisive tendencies of current scientific movements are retarding progress.

The key factor in the current controversy about the nonphysical nature of the human spirit really hinges on a definition: Just what do we mean by physical structure and what are its properties? We must define the physical before we can talk rationally about the nonphysical. A brief summary of a few principles of current physical science is prerequisite to further consideration of the meaning of the word "physical" in biological and psychological contexts.

PHYSICAL SCIENCE OF TODAY

The history of ideas about the nature of the physical world records three revolutionary periods. The Greek period, typified by Aristotle and Euclid, dominated formal thinking for two thousand years and is still influential. Beginning in the middle ages a second revolutionary period culminated with the Newtonian system of mechanics. The third period began with the twentieth century, as exemplified in relativity and quantum mechanics. It is still in its infancy and its findings are more revolutionary than any of the preceding.

The fundamental conceptions of physics are in flux. New methods have revealed new facts which require new principles, and some principles formerly regarded as axiomatic are now suspect. The absolutes of earlier times are now treated relativistically, with radical reorganization of the science of mechanics.

To put it briefly, natural science today regards our cosmos as a stupendous mechanism (physicalism) composed of innumerable subsidiary mechanisms, all bound together in accordance with lawfully ordered principles. We do not adequately understand any thing or any event until we know the mechanism that produces it and the principles in accordance with which it operates. A mechanism is defined by D'Arcy Thompson (7) as ". . . whatsoever checks or controls, and guides into determinate paths, the workings of energy" (p. 291). Accepting this definition, the science of mechanics deals with energy and the "whatsoever" that controls its workings.

The mechanism makes some specific kind of product and the nature of the product is the crucial issue in its organization. This product may be material arranged in a different way or place, or energy in a different pattern of manifestation. Or it may be matter transformed into energy or energy into matter, for these are known to be interconvertible in quantitatively measurable relations (Einstein's conversion equation, $E = Mc^2$). In view of this last point the distinction between matter and energy becomes rather fuzzy, and any manifestation of an energy change is a physical event.

The belief now current among physicists is that the various kinds of atoms are relatively stable and different patterns of energy. There are no different kinds of energy. The so-called thermal, electrical, and other energies are different patterns of manifestation of one common measurable quantity which in our ignorance we call energy. Nature as a whole is process. There is nothing static about it anywhere.

In a natural mechanism the materials and energies used may come from a wide field and the product made is in turn delivered to the surroundings. It follows that the active structure, the machine itself, must be so defined as to include the entire field with which it has transactional relations. The mistaken popular notion that a machine is a passive inert structure operated by an outside agent is derived from the artificial machine with a human operator. It is true that a lathe in a machine shop has no value as mechanism without the operator who starts it, stops it, and controls its action. This means that, from the operational standpoint, the operator must be regarded as an integral part of the mechanism. In natural mechanisms it is more evident that the operator is inside the apparatus. The apparatus itself is the operator; that is to say, the machine operates itself. No natural mechanism needs a djinn or any external operator to run it or tell it what to do. This, we believe, is as true of a man as of a solar system or a volcano.

Classical physics as formulated in Newtonian mechanics deals with inert solid particles of matter which differ in mass and are pushed about by forces acting upon them. These forces are considered to be manifestations of energy, and the movements are measurable in arbitrary units of space and time. Twentieth century science, on the contrary, finds that matter and energy are different manifestations of the same unknown something and that in some domains of physics space and time in the objective world must be treated relativistically with reference to each other. In subatomic physics they cannot be measured in arbitrary units, with the observer as a fixed point of reference.

A physical mechanism is defined dynamically. It may make some particular product, repetitively, because it is so organized as to do this by virtue of this organization. But if the organization changes its pattern so as to deliver a different kind of product, a factor is introduced which may properly be called creative. Even a repetitive performance like that of some particular chemical reaction exhibits the property of transforming a pattern of material or energy into a different pattern, and this capacity is the source from which creativity, as here defined, is derived. Creation does not imply that something is made out of nothing. The scientific problem is to discover the laws in accordance with which these changes take place. Since the operating forces are manifestations of energy, it is evident that energy as such has creative efficiency. In other words, its activities are not stereotyped in rigidly predetermined patterns. These patterns are constantly changing by conversion as they interact with one another, and this capacity for change is responsible for cosmic and organic evolution and for the orderly processes of growth of living individuals. This perhaps is what Whitehead had in mind when he said, "creativity is ultimate"; that is to say, it is something which science cannot explain or analyze further.

This intrinsic creativity of physical processes is not a metaphysical speculation. It is an observed fact, and if it were not so our cosmos would now be in a state of homogeneous equilibrium at the lowest level of organization. In view of these dynamic and creative properties of the physical universe, and of our incomplete knowledge of the laws in accordance with which the successive changes take place, we can set no limits to the kinds of products that natural mechanisms (physicalisms) can make. If now we wish to push further into the domain of psychophysics, account must be taken of this change in the climate of science which has come within the memory of men now living.

The modern mechanist is not a materialist in the classical sense. We grant that he deals only with physical structures and their operations. The operating body is always a mobile structure, which may be relatively stable as we see it in a mountain or a cat, or it may be a specific pattern of energy manifestation in which no particulate matter is recognizable, as in a ray of light. Usually it is a combination of both, as in a volcano or a thinking man. In current science by "mechanistic" we mean physicalistic, and these words now have meanings which are quite different from those of fifty years ago, although the latter are still current in both common speech and scientific literature.

Subsequent references to relativity and quantum mechanics do not carry the implication that the psychobiologist thinks that these principles, as now formulated in mathematical physics, are adequate to give a satisfactory explanation of the mind-body relationship. The most that can be claimed for them in this connection is that they may be steps toward the discovery of the still unknown principles that satisfy the requirements of the problem.

THE NATURAL HISTORY OF THE SPIRIT

My understanding of current scientific thought about some of the properties of the natural world has just been outlined. It is physicalistic throughout. Let us see now where the opposing psychologies mentioned at the beginning of this essay stand in the light of this evidence.

Many years ago I said (2) that introspection is as manifestly a biological method as is reflexology, and this principle needs re-emphasis today. Conscious experience influences conduct. It is therefore a factor of behavior. We may go further and say with Edwin G. Boring that a conscious experience is itself a behavior, a bodily act which may be observed introspectively just as we observe other kinds of behavior. Introspection, like the other methods of observation, is, of course, subject to the hazard of erroneous interpretation. And again it is timely to insist that the dualistic theories merely pile mystery upon mystery and only confuse the issues without explaining anything.

Because man is an animal and because his "spiritual" capacities are demonstrably vital processes, psychology is necessarily articulated with biology; but it does not follow from this that the laws of conscious experience are identical with those of the things of which we have experience. It is, in fact, clear that they are not. Each level of organization has its own distinctive properties, some of which cannot be reduced to those of lower levels (3).

The mind-body problem will never be solved by ignoring the troublesome factors, either those of spirit or of matter. The inquiry cannot be limited to either the conscious or the unconscious factors, because what we are looking for is the relation between the two. It does no good to try to evade the issue as a "pseudoproblem." Traditional materialism (the "crude" variety) and classical spiritism (or, more reputably, "idealism") both involve neglect of a vast wealth of human experience, including common sense and refined scientific knowledge. We cannot choose between materialism and spiritualism. We must have both.

The attempts so far made to find a

suitable formula of psychophysics have led into blind alleys. The chief difficulty arises from the fact that we have no common measure for objectively known dimensions and the relations observed in the process of knowing them. If ever a suitable formula for this relationship is discovered, it will probably be stated in some kind of relativistic terms, and it may dispense with numerical units altogether. Mathematics and symbolic logic are now developing methods of dealing with situations devoid of any quantitative factors, and this may point the way toward hitherto unexplored fields of inquiry in psychobiology.

The human brain controls many of the internal bodily activities and by far the larger part of overt behavior. Its structure is inconceivably complicated and much of the detail is still obscure. There is on record, however, a vast amount of knowledge about nervous structure and the laws of its operation. The products delivered by this mechanism are of many kinds, including growth of the structure itself, a variety of chemical reactions and associated changes in electrical potential, and the excitation of mass movements of the body. These have been thoroughly studied and described in physiological terms.

There is another unique property of some of these operations that cannot be studied by the objective methods of anatomy and physiology, because it is strictly private and can be recognized only subjectively by the person who is operating. Fortunately, we have a mechanism by which this personal awareness can be made public. Ideas can be translated into symbols—words and others—the utterance of which evokes similar symbols in other people. By this indirection subjective experience can be objectified and so communicated. This enables us to develop a legitimate science of introspective psychology. We do not know the mechanics of this process of symbolizing or the physical laws of its operation, but we have an interesting analogy in the electronic computing machines which perform similar operations with a quite different mechanism. Doubtless there are some common principles in these two kinds of operations.

There are no conscious factors, so far as we know, in most physiological operations; but we know as surely as we know anything in science that all mental acts are vital functions and we know where the organs are that perform them. We know also that the distinctively human types of higher mental functions are added to pre-existing physiological functions in the course of personal and evolutionary development.

Although we do not yet know how awareness emerges within the metabolism of brain tissue, we know that it does so and that it is as truly a physical process as is the transmission of light through apparently empty space. We do not know all about the mechanics of light either, but long ago Helios was banished from science because he could not explain anything. When a weight is lifted by muscular action the cause of the movement is not a disembodied contractility. It is contracting muscle. When a problem in mental arithmetic is solved the operating agent is not a nonphysical mind which activates an inert structural mechanism. It is a living brain engaged in thinking. There is no specific kind of mental energy. Mental work is bodily work, and it is the body that gets tired when we do it. Mind does not move matter. It is "minding" matter that does "mental" work just as it is contracting muscle that does "physical" work. A thought is not a product made by a mechanism, the way bile is made by the liver. It is the operation of the mechanism, an act, not a material or any other kind of an object. *To split the function from the organ that performs it is scientifically as absurd as is the separation of the properties of any other thing from the object that manifests them.*

The claim of the parapsychologists that psi phenomena in general involve "extrasensory perception," presumably through nonphysical agencies, calls for a critical appraisement of current studies of the mechanics of perception. This inquiry brings out very clearly the fundamental difference between the relations of space and time factors in the acquisition of perceptual knowledge of the objective world and these relations in the higher rational use of this knowledge.

Every perception of a physical object or event and every consciously controlled movement of the body involves a transactional relation of some sort between the "spiritual" and the "physical." In all these activities the conscious and the unconscious factors are inextricably interwoven. The first can be observed only introspectively, the other again introspectively but only by an indirection, by the observer becoming aware of some objective (unconscious) physical events. The act of perception involves the conversion of "physical" processes into "mental" processes, and conversely a voluntary act requires the conversion of a conscious purpose into movement of the physical body. The nub of the mind-body problem lies in the nature of these conversions.

The significant factors of such conversions can be isolated by experimental analysis more readily in the study of perception than of any other psychological processes. We have biological evidence that the internal integrative and regulatory functions of an animal body must be sharply distinguished from the analytic functions of sensorimotor type which are concerned with the adjustment of the body to external things and events. The physical principles of integration are fundamentally different from those of the analytic functions, and some of the former can be described only in relativistic terms.

The most complicated integrative apparatus known is in the human brain, and the act of knowing is an integrative process which gives us two kinds of knowledge which Sellars (6) has called "perceptual knowledge" and "conceptual knowledge." The former gives us all the information we have about the spatial, temporal, and energic properties of the objective world. Newtonian mechanics was developed on the basis of this knowledge. This mechanics is of necessity framed in behavioral space and time. Conceptual knowledge comprises those higher rational and emotive functions which cannot be quantified in the numerical units of space, time, mass, and energy of Newtonian mechanics.

This distinction is fundamental. It is explained, in part at least, by the nature of the act of perception and of the apparatus employed. Its neglect accounts for the failure of the earlier attempts to find acceptable laws of psychophysics. This theme cannot be elaborated here. Enough has been done to show that this lead points to a promising field worthy of further investigation.

PARAPSYCHOLOGY

The intimate relationship between objective things and events and our knowledge of them is so commonplace that it is generally uncritically accepted. But the manifestations of hypnotism, telepathy, and the other so-called psi phenomena are so unusual, and in some instances so bizarre, that there is insistent demand that they be either explained or explained away as spurious. The latter has been tried repeatedly without convincing proof.

These phenomena have been under investigation by competent scientists for more than seventy-five years. The workers in this field of "parapsychology" who are now most active are advocates of dualistic theories which have been given wide publicity, but the majority of other scientists question the significance of their findings. This skepticism is due in part to the fact that this field of inquiry was so long contaminated with fraud and delusion that the prejudice against it is still strong. The prejudice has not been mitigated by the fantastic claims made in some of the recent literature. Nevertheless, these are real phenomena which merit critical study by every available method.

It should be recognized and emphasized that the mysteries associated with psi phenomena are of the same order, and no different in principle, as those presented by every mental and intentionally controlled act whatsoever. If parapsychology is to win acceptance as legitimate natural science, it must be fitted into the same general frame of reference as the other branches of psychology and of natural science as a whole. "Parapsychology" is a misnomer, for psi phenomena are either psychological or they are not. If not, what are they?

In psi phenomena the unconscious factors clearly play the critical role. These can be studied only by the methods of objective science and these methods have so far proved inadequate. The same is true for many other unsolved scientific problems. So when Dr. Rhine writes (5, p. 300), "We know by this time that we cannot use a physical theory to explain psi," we freely grant that no satisfactory physical theory has yet been found; but that is far from proof that it cannot be found. In another passage (4, p. 62) he has insisted that the physics of tomorrow is irrelevant here; we must think in terms of present knowledge. But this is exactly what he does not do. What he does do is to base his argument on nineteenth century physics. As for the physics of tomorrow, which Rhine says is a pointless speculation, we should remember, as Walker (8) points out, that our most valuable instruments of scientific progress are the working hypotheses which look forward by extrapolation from present knowledge in all promising directions. Without these prescient excursions into the unknown, growth would be arrested and science would die of senile sclerosis.

We are surprised also to read that the claim that human mentation as a vital function is "a metaphysical assumption" or "mere conjecture," and that psychologists have the vague notion that mind and body "are somehow fundamentally unified on some complex but wholly unknown physical basis," and that "this half-formulated materialism is taken for granted; it has not been subjected to experiment" (5, p. 197). The fact is that it would be possible to cite literally thousands of experiments devoted explicitly to this problem—experiments made by the most competent neurologists, physiologists, and psychologists we have. These experiments are of many kinds, using a wide variety of methods.

The most instructive of these experiments are based on the fact that all nervous action is accompanied by changes in electrical potential that can be localized and accurately measured. Using the same technique, a particular kind of mental work can be shown to be accompanied by similar changes in potential which are localized in specific regions of the body. The "brain waves" recorded by the oscillograph of thousands of conscious persons have been studied under a great variety of conditions, and the effects of various kinds of mental experience are clearly seen in this permanent record of the related changes in electrical potential. This is only one among many other methods which have been used to prove that we think with our bodies, and with different parts of the body for different kinds of thinking.

THE SPIRITUAL LIFE

It is possible to mention here only a few samples of the evidence which justifies the conviction that man's spiritual life, in the ordinary vernacular meaning of that term, is a real and significant component of his natural life. For the biologist this means that all mental acts of whatever kind are as truly vital processes as are nervous conduction and muscular movement. This gives us a unitary system of scientifically acceptable principles for everything within the range of human experience.

A man has native capacities for spiritual culture which set him at the highest level of integrative and creative efficiency yet reached in cosmic evolution so far as we know. This physicalistic conception of human nature, which has been accused of degrading the dignity of man and destroying his spiritual values, really points the way to the most efficient measures for enhancing these most precious human treasures. By setting the spiritual life in vital operational relationship with all other domains of science and of human endeavor, we get the benefit of all the resources of the sciences to reinforce and guide into appropriate channels the efforts of the other agencies of spiritual culture—education, philosophy, art, religion, and all the rest. This is a magnificent achievement.

REFERENCES

1. GIBBS, F. A. The most important thing. *Amer. J. publ. Hlth,* 1951, 41, 1503–1508.
2. HERRICK, C. J. Introspection as a biological method. *J. Phil.,* 1915, 12, 543–551.
3. HERRICK, C. J. A biological survey of integrative levels. In R. W. Sellars *et al.* (Eds.), *Philosophy for the future.* New York: Macmillan, 1949. Pp. 222–242.
4. RHINE, J. B. *The reach of the mind.* New York: William Sloane, 1947.
5. RHINE, J. B. *New world of the mind.* New York: William Sloane, 1953.
6. SELLARS, R. W. An analytic approach to the mind-body problem. *Phil. Rev.,* 1938, 47, 461–487.
7. THOMPSON, D. W. *On growth and form.* New York: Macmillan, 1942.
8. WALKER, R. Parapsychology and dualism. *Sci. Mon.,* 1954, 44, 1–9.

(Received October 14, 1954)

Psychological Review
Vol. 62, No. 5, 1955

THE STIMULUS CONDITIONS WHICH FOLLOW LEARNED RESPONSES

CHARLES C. PERKINS, JR.

Kent State University

The further development of a reinforcement theory of learning, or, for that matter, of any other general learning theory, would seem to require a systematic analysis of the different procedures used in the experimental investigation of basic learning processes. The present paper will be limited to a part of this task, the analysis of the stimulus conditions which immediately follow occurrences of the response acquired or extinguished.

The approach taken here is to assume that any stimulus-response connection which becomes stronger is followed by a more reinforcing stimulus situation than alternate responses which tend to drop out. The frame of reference of this paper resembles that of Meehl (11) in his evaluation of the law of effect.

An analysis of the stimulus situations which follow the alternative responses in a particular context is assumed to disclose the relative reinforcing properties of these stimulus situations. Once the relative reinforcing properties have been determined by the observation of learning, they can be used to predict learning of other responses. This approach, one typical of explanations of trial-and-error learning, will be applied to other "types" of learning as well. I will then show that the results of the analysis seem to require a somewhat different treatment of conditioning from that represented by typical reinforcement theory (7, 12), contiguity theory (2, 5), or two-factor theory (13, 16). The analysis to be made below seems to clarify the distinction between trial-and-error learning and conditioning at the same time as it implies that common principles underlie all learning.

In the final sections I will (*a*) show how the analysis suggests an explanation of the development of observing responses under conditions which would not lead one to predict their acquisition from other viewpoints; and (*b*) suggest several direct experimental tests of the interpretation of learning outlined in this paper.

Since an analysis is more meaningful if terms are explicitly defined, we now turn to brief definitions of some of the terms and symbols used in this paper.

DEFINITIONS AND SYMBOLS

As the term is used in this paper, a *stimulus* is any situational variable which, if appropriate training were given, could come to be followed by a response which would not occur in the absence of this stimulus. Any stimulus which is an environmental condition present at the moment being described is an *external stimulus*.

An *internal stimulus* is the stimulus aftereffect of (*a*) an environmental condition which was present earlier (whether or not it remains at the moment in question), (*b*) a preceding response, or (*c*) a deprivation schedule.

A *situation* is the stimulus complex at a given moment in so far as it is specified by a report of the experimental procedure. Since no description includes all observations or measures that could be made, a situation (like a stimulus) is a generic term referring to a class of situations. However, unless otherwise indicated, it will be assumed that reports of experimental procedures are complete enough to indicate (at least by implication) any systematic difference between situations which precede a correlated difference in subsequent events.

It should be noted that the above definition of a stimulus includes stimuli defined as changes in the environment. However, unlike the latter definition, ours urges an inventory of all situational variables. Such a listing of situational variables appears to facilitate the determination of response-evoking conditions as well as conditions affecting reinforcement.

The term *response* will be used to include the nonoccurrence of an activity as well as the occurrence of an activity. This usage seems preferable because "resting," or not making a given response, appears to be learned in the same way that an active response is acquired.

Our use of the term *reinforcement* is best explained in conjunction with a description of trial-and-error learning, to which we now turn.

TRIAL-AND-ERROR LEARNING

The basic working hypothesis of this paper is the assumption that whenever two responses occur to the same stimulus situation, and the tendency for one (R_A) to occur gains strength at the expense of the tendency for the other (R_B) to occur, the stimulus situation which follows the first (X_A) is more reinforcing than the situation which follows the second (X_B).[1]

Actually, to avoid circularity, X_A and X_B must refer to classes of situations. Perhaps the classes of situation most frequently contrasted in learning experiments are a context which includes hunger (X_c) and food (X_a), on the one hand, and the same context (X_c) and no food (X_b), on the other.[2] Hunger is, of course, just a shorthand way of labeling the aftereffect of a period of food deprivation.

Strictly speaking, the stimulus aftereffects of R_A and R_B cannot be identical, and therefore situations X_A and X_B cannot be identical either. However, the contexts following R_A and R_B may be equivalent; i.e., stimuli may have the same relative reinforcing properties in either context. For example, either right or left turns in a T maze are ordinarily learned about equally rapidly when reinforced by food or any one of a number of other reinforcing stimuli. Thus the context of goal box following a left turn is equivalent to goal box following a right turn.

When two situations are equivalent except that the first includes X_a and the second X_b, and R_A increases relative to R_B, then X_a (a stimulus, not a situation) may be said to be more reinforcing than X_b. However, it should be recognized that this does not necessarily imply that X_a would be more reinforcing than X_b in another context. Studies of drive discrimination (6, 9) provide perhaps the best known experimental results which demonstrate clearly that the more reinforcing stimulus (water as more reinforcing than food) in one situation (which includes water deprivation) is less reinforcing in another context (food deprivation). However, if we recognize the implied specificity of the context it is meaningful to speak of reinforcing stimuli as well as reinforcing situations.

---

[1] A further advantage of our definitions of response and of reinforcement is that the question of whether punishment weakens preceding response tendencies, or only leads to the strengthening of competing response tendencies, becomes meaningless. If R_A and R_B are two mutually exclusive response categories which include all possible responses, there is no way of discriminating between the strengthening of R_A and the weakening of R_B.

[2] Capitalized subscripts refer to situations, while stimulus components are designated by uncapitalized letters. Similarly, S_A or S_a refer to the situation or stimulus which precedes R_A, while X_A and X_a designate events which follow R_A.

When the expression "reinforcing stimulus" is used, it ordinarily implies that the situation plus this stimulus is more reinforcing than the situation without this stimulus. The implied description of situation minus stimulus is usually unambiguous. Similarly, a noxious stimulus ordinarily implies that the situation plus this stimulus is less reinforcing than the situation without this stimulus.

The above example of learning in a single choice-point maze is ordinarily considered a typical case of trial-and-error learning, and our analysis does not differ markedly from the usual explanation. However, after an examination of several special, atypical cases of "trial-and-error" learning we shall show that the same sort of analysis is quite adequate for an explanation of Pavlovian conditioning.

One technique for establishing a tendency for a striped-muscle response to occur in a given situation where it originally had little or no tendency to occur is the differentiation of a response (3). This technique consists first of reinforcing any responses which occur and have some similarity to the desired response; then, after the tendency to make similar responses is sufficiently strong, reinforcement is gradually restricted to those responses most similar to the response the experimenter wishes to establish. Later, after the desired response has occurred, reinforcement of all other responses is withheld.

This technique has probably been most fully described in connection with training pigeons to peck a key. However, it is a common procedure as part of preliminary training in other situations, and would appear to be applicable to the attaching of almost any striped-muscle response which the subject is capable of making to practically any stimulus situation.

In differentiation of a response, as in typical trial-and-error learning, the reinforcing stimulus is either presented or withheld; in other words, *all-or-none reinforcement* is used. However, *varied reinforcement*—Hull (8) has used the more descriptive but more cumbersome expression "correlated reinforcement intensities"—is another procedure which can also lead to the acquisition of a response which would not otherwise occur and thus could not be reinforced. In target shooting and many other motor skills the degree of reinforcement depends on the accuracy of the response. Therefore, the "better" or more accurate responses are always reinforced relative to the "worse" or less accurate ones.[3]

Use of varied reinforcement to establish new striped-muscle responses in animal subjects probably would not be a practical experimental procedure. The difficulty of providing a sufficiently distinctive dimension of reinforcing stimuli could not be easily overcome. This is perhaps the reason Hull applied correlated reinforcement intensities only to examples in which human subjects acquired responses which were strengthened by secondary reinforcement.

Three types of trial-and-error learning have been distinguished above. They are: (*a*) the most typical case, all-or-none reinforcement with the same class of responses reinforced throughout; (*b*) all-or-none reinforcement with gradual restriction of the responses reinforced; and (*c*) varied reinforcement, where the degree of reinforcement varies depending on the exact nature of the response. Since "new" responses can be acquired by either of the latter two procedures, the claim of some writers (10) that

---

[3] Actually, if level of aspiration is emphasized, learning of many skills can be adequately treated in terms of all-or-none reinforcement, in which case the learning would be a special instance of differentiation of the response. Probably both changes in level of aspiration and varied reinforcement are ordinarily involved in the acquisition of such skills by human subjects.

"new" responses are never acquired by trial-and-error learning does not fit the facts.

PAVLOVIAN CONDITIONING

It is the contention of the writer that the major difference between trial-and-error learning and Pavlovian conditioning is the nature of the reinforcing stimulus.

Since our basic premise is that responses followed by a more reinforcing stimulus situation are strengthened at the expense of those followed by a less reinforcing situation, we must look for the difference between the situation which follows marked salivation and the situation which follows little or no salivation in a Pavlovian conditioning experiment. Apparently there are no external stimuli which follow the salivary response to the CS but do not follow presentation of the CS and absence of the salivary response. Thus, the learning must depend upon reinforcement by an internal stimulus.

The only stimulus which systematically differentiates between the conditions immediately after presentation of the CS on trials on which a conditioned salivary response occurs and trials on which the CR does not occur is the stimulus aftereffect of salivation, or saliva-in-the-mouth.[4] Therefore, the occurrence of salivary conditioning when the US is food implies that a situation consisting of context, food, and saliva is more reinforcing than a situation consisting of context, food, and absence of saliva. The reinforcing stimulus, the stimulus component which distinguishes between the two situations, is saliva—an internal stimulus, not a part of the external environment.

Extinction of the salivary response when the US no longer follows the CS implies that a situation made up of context, absence of food, and absence of saliva is more reinforcing than a situation consisting of context, absence of food, and presence of saliva. Since salivary conditioning appears to be a general phenomenon, food plus saliva must be more reinforcing than food without saliva in a wide variety of contexts. Similarly, it appears that in many contexts the compound stimulus absence of saliva and absence of food is more reinforcing than presence of saliva and absence of food. The generality of these phenomena permits prediction of learning in a new context. Thus, our treatment is not circular.

A possible objection to the above treatment of Pavlovian conditioning is that the conditioned response has no initial tendency to occur and therefore cannot be reinforced. However, it can be demonstrated that such an objection does not apply if the following points are recognized: (*a*) some salivation occurs normally even if no food or "signals" of forthcoming food are present; (*b*) presumably there are fluctuations in rate of salivation from one trial to the next; and (*c*) varied reinforcement is involved in salivary conditioning.

The relation between rate of salivation and the reinforcing properties of the situation which follows is presumably similar to that shown by the solid line in Fig. 1. On the early conditioning trials the salivary response will be less than x (Fig. 1), but will vary between 0 and x (from points *a* and *b* above). During these trials the greater the amount of salivation to the CS the more reinforcing the situation which follows. A positive relationship between salivation and degree of subsequent reinforcement exists until rate of salivation surpasses *Opt.*, the rate of salivation which is followed by maximum reinforcement. Thus, rate of salivation will gradually increase until the mean rate is at *Opt.*, and since further increase in salivation is followed by a less reinforcing situation, rate of salivation will stabilize at point *Opt.*

---

[4] For convenience, saliva or saliva-in-the-mouth will be used interchangeably with the stimulus aftereffect of salivation, even though there are probably also other stimulus aftereffects of salivation. It should be noted, however, that the physiological nature of stimulus aftereffects of an event (e.g., saliva in the mouth) need not be specified for postulation of stimulus aftereffects. The expression "saliva in the mouth" is used here on the assumption that it will facilitate the exposition for physiologically minded readers.

The dotted line in Fig. 1 represents the reinforcing properties of stimulus situations which follow various rates of salivation when no food is presented, as on extinction trials. The process of extinction may be explained in the same manner as was conditioning.

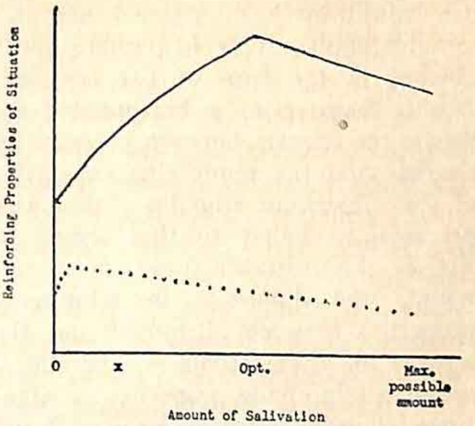


FIG. 1. The hypothetical relation between amount of salivation and the relative reinforcing properties of the stimulus situation which follows the CS on food-reinforced trials (solid line) and on extinction trials (dotted line).

This treatment of learning implies that a conditioned response is a preparatory response. This interpretation is strongly supported by recent work (1, 4, 14, 18) which rather clearly indicates that effective conditioning occurs only when the CS precedes the US. Instances of apparent backward or strictly simultaneous conditioning can all be explained as special cases of pseudo conditioning. Pseudo conditioning in turn appears to be only a special instance of the operation of stimulus generalization (19).

TYPES OF RESPONSE MEASURES

The type of response measured is closely related to the type of reinforcement. As our treatment of salivary conditioning illustrates, smooth-muscle responses would seem to be acquired only through varied reinforcement by internal stimuli.[5]

Striped-muscle responses (and the responses of not engaging in certain striped-muscle activities) are typically reinforced by external stimuli as external stimuli are defined in this paper. However, there are some exceptions—as, for example, the response of not biting the tongue which is reinforced relative to biting the tongue.

A particularly important class of internal stimuli which determine the extent to which striped-muscle responses are reinforced is the stimulus aftereffects of activity or "fatigue" effects. Specifically, this class of stimuli provides greater reinforcement for inactivity than for activity.

If the response acquired is bar depression or a certain turn in a single choice-point maze containing food on but one side, and if only the occurrence or nonoccurrence of the response is recorded, internal stimuli (such as fatigue effects) need not ordinarily be considered. In such studies all occurrences of, say, bar pressing are followed by a more reinforcing state of affairs than any occurrences of not pressing the bar.

If pressure on the bar or running speed is the response measure, the relative reinforcing property of the situation following a particular response is determined by both internal and external stimuli. This is perhaps most obvious in the case of bar pressing. During learning of a bar-pressing response the relation between pressure on the bar and the reinforcing properties of the subsequent stimulus situation is presumably similar to that shown in Fig. 2. The distance from *a* to *b* represents the difference in reinforcing properties between click-food and the absence of these stimuli. The differences in reinforcing properties of situations following responses between 0 and *x*, and between *x* and *m*, result from kinesthetic and proprioceptive stimuli.

---

[5] Smith's ingenious suggestion (17) that smooth-muscle conditioning is an artifact resulting from trial-and-error learning of striped-muscle responses and unlearned bonds between stimuli from striped muscles and smooth-muscle responses is essentially a physiological theory which has no necessary relevance to the subject of the present paper. However, Smith's interpretation fits in very harmoniously with the behavioral principles advocated here.

When running speed is the response studied, a similar analysis can be made. However, a detailed analysis of this type of learning is a good bit more complex and would require consideration of secondary reinforcement and/or anticipatory responses.

The definition of (relative) reinforcement used here has certain disadvantages which do not characterize an absolute (e.g., drive-reduction) definition. In particular, knowledge of the reinforcing properties of a stimulus in one context is not enough to predict the reinforcing properties of this stimulus in another context. However, even a drive-reduction theory implies as much; the drive (e.g., deprivation schedule) and goal object together are required for reinforcement. Thus, unless drive is specified in terms of physiological activity rather than antecedent environmental conditions, drive reduction implies that the reinforcing properties of a stimulus depend on contextual stimuli. Unless it has acquired secondary reinforcing properties, food is a reinforcing stimulus only if there is also a hunger stimulus.

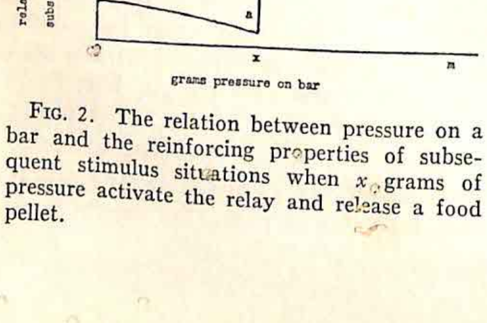


FIG. 2. The relation between pressure on a bar and the reinforcing properties of subsequent stimulus situations when *x* grams of pressure activate the relay and release a food pellet.

It thus appears that, for a molar behavior theory, stimulus patterning is an inescapable fact which affects not only the discriminative (perceptual) properties of situations but also their reinforcing properties.

THE ACQUISITION OF OBSERVING RESPONSES

Wyckoff (20) developed an ingenious technique which made possible the measurement of observing responses (R_o). His apparatus was a pigeon "Skinner box" in which there was a treadle as well as a key which was attached to the food-release mechanism. The apparatus was so wired that when the subject made the R_o of stepping on the treadle the key became either red or green.[6] When the subject was not standing on the treadle the key was white. Pecking at the red key was periodically reinforced by food; pecking the green key was never followed by food. Presentation of food was unrelated to the occurrence of R_o so that when R_o did not occur food would sometimes follow pecks to the white. The subjects showed a clear increase in R_o under these conditions, and a marked decrease when food presentation was no longer related to the color of the key. Wyckoff postulated that "information" has reinforcing properties and suggested, without elaboration, that secondary reinforcement is probably involved. Thus, in effect, he described his findings but did not derive them from other principles.

---

[6] Strictly speaking, R_o is not an observing response in this study or in Prokasy's. Following Wyckoff's usage, we use R_o to refer to any response which provides discriminative (positive and negative) stimuli which are not provided unless R_o occurs.

Prokasy (15) ran rats in a single choice-point maze where either turn led to food on half the trials. All turns were followed by a 30-sec. delay before entrance to the end box. Either turn led to a black delay and goal box on half the trials, and a white one on the other half of the trials. On the consistent side one color (black) delay box was followed by access to food 100 per cent of the time, while the other (white) never led to food. On the inconsistent side either color delay box was followed by access to food on half the trials. Thus, R_o represents selection of the consistent side.

An increase in percentage of free turns to the consistent side was acquired by between 9 and 11 of the 12 subjects depending on the particular measure used. It should be noted that the two sides did not differ with respect to the amount of food received or the color of the goal and delay boxes. The acquired preference could only have resulted from the relation between color of delay box and food presentation.

Prokasy's findings follow logically from an analysis of the reinforcing stimuli following occurrence of a turn to either side.

Table 1 contrasts the stimuli which follow choice of the consistent side with those following choice of the inconsistent side, on food-reinforced and on non–food-reinforced trials, subsequent to the acquisition of anticipatory salivary responses.

Since each of the situations following responses to the consistent side is more reinforcing than the corresponding situation (same food condition) following a choice to the inconsistent side, the former choice will be acquired. A similar analysis can be used to explain Wyckoff's results.

DIRECT EXPERIMENTAL TESTS

The viewpoint outlined in this paper can most easily be subjected to direct experimental tests in connection with Pavlovian conditioning. While there are obvious technical difficulties, surgical techniques which would completely prevent salivary secretion from entering the mouth should lead in time to a decrease in the rate of conditioning and extinction, and perhaps to the eventual absence of any relation between salivation and stimulating conditions. Drive-reduction, contiguity, and two-factor theories would not seem to lead so clearly to such a prediction. However, this most certainly would not be a crucial experiment, since any general principle can be made consistent with any single experimental result.

If technical problems were overcome, several further experiments could be conducted involving the clamping of the salivary ducts so that the flow of saliva could be prevented whenever the experimenter wished. For example, the ducts could be blocked whenever the subject was placed in a highly distinctive locus and at no other time. If the subject had been frequently fed in this locus, we might expect that salivation would no longer accompany eating here,

TABLE 1

THE STIMULI SUBSEQUENT TO CHOICE OF EITHER SIDE IN PROKASY'S EXPERIMENT

Goal-Box Condition	Stimuli Following a Response to the Consistent Side (Correct "Anticipation" Possible)	Stimuli Following a Response to the Inconsistent Side (Correct "Anticipation" Impossible)
Food*	Food plus optimum (much) salivation	Food plus less than optimum salivation
No food*	No food plus optimum (very little) salivation	Food plus more than optimum salivation

* Food followed either response 50 per cent of the time.

as saliva would not enter the mouth, but instead salivary gland activity might inflate the gland to the point where it would be "painful." Presumably, the subject would eventually learn not to salivate while eating in the locus in which the ducts were blocked, but still continue to salivate while eating elsewhere.

Another possible experiment would involve a system for temporary storage of the subject's saliva in a container, and the injection of this saliva into the subject's mouth when a particular striped-muscle response is made. For example, it should then be possible to train a dog to stretch his right foreleg before and during eating in order to inject saliva into his mouth.

CONCLUSION

The emphasis on relative reinforcement suggested in this paper seems to have several distinct advantages, most notably that (*a*) it allows for a treatment of all learning by the same set of principles, without involving the complications which a drive-reduction theory necessitates;[7] and (*b*) it permits derivations concerning the acquisition of observing responses which are not instrumental in attaining a goal object from principles already demonstrated experimentally.

REFERENCES

1. BORASIO, G. An experimental investigation of backward, forward, and strictly simultaneous conditioning of fear. Unpublished master's thesis, Kent State Univer., 1952.
2. ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
3. FERSTER, C. B. The use of the free operant in the analysis of behavior. *Psychol. Bull.*, 1953, **50**, 263–274.
4. FITZWATER, M. E., & REISMAN, M. N. Comparisons of forward, simultaneous, backward, and pseudo-conditioning. *J. exp. Psychol.*, 1952, **44**, 211–214.
5. GUTHRIE, E. R. *The psychology of learning.* New York: Harper, 1935.
6. HULL, C. L. Differential habituation to internal stimuli in the albino rat. *J. comp. Psychol.*, 1933, **16**, 255–273.
7. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
8. HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
9. LEEPER, R. The role of motivation in learning: a study of the phenomenon of differential motivational control of the utilization of habits. *J. genet. Psychol.*, 1935, **46**, 3–40.
10. MAIER, N. R. F., & SCHNEIRLA, T. C. Mechanisms in conditioning. *Psychol. Rev.*, 1942, **49**, 117–134.
11. MEEHL, P. E. On the circularity of the law of effect. *Psychol. Bull.*, 1950, **47**, 52–75.
12. MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.
13. MOWRER, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.
14. MOWRER, O. H., & AIKEN, E. G. Contiguity vs. drive-reduction in conditioned fear: temporal variations in conditioned and unconditioned stimulus. *Amer. J. Psychol.*, 1954, **67**, 26–38.
15. PROKASY, W. F., JR. The relative effects of two reinforcement procedures on a choice response. Unpublished master's thesis, Kent State Univer., 1954.
16. SKINNER, B. F. *The behavior of organisms.* New York: D. Appleton-Century, 1938.
17. SMITH, K. Conditioning as an artifact. *Psychol. Rev.*, 1954, **61**, 217–225.
18. SPOONER, A., & KELLOGG, W. N. The backward conditioning curve. *Amer. J. Psychol.*, 1947, **60**, 321–334.
19. WICKENS, D. D., & WICKENS, C. Some factors related to pseudo-conditioning. *J. exp. Psychol.*, 1942, **31**, 518–526.
20. WYCKOFF, L. B., JR. The role of observing responses in discrimination learning. *Psychol. Rev.*, 1952, **59**, 431–442.

(Received November 15, 1954)

---

[7] This should not be taken as implying that the viewpoint expounded here is necessarily incompatible with a drive-reduction theory. In fact, if the viewpoint presented here is extended, it would seem to imply that all increases in response tendency result from "drive" reduction. However, the approach suggested in this paper seems more parsimonious and more easily related to experimental findings.

Psychological Review
Vol. 62, No. 5, 1955

CONSCIOUSNESS, BEHAVIOR, AND PERSONALITY

JOSEPH NUTTIN

University of Louvain

The concept of consciousness still remains associated with one of the most traumatic experiences in the life history of scientific psychology, i.e., introspectionism. The wound has not yet healed; consciousness still means for many people a menace to behavioral science. It would not be justified, therefore, to run the risk of arousing emotional reactions by reintroducing the subject if there were no hope that a revocation and a restatement of the problem in the present situation might entail some therapeutic effect and eventually help psychology in its development to greater maturity. Maturity, in fact, means getting in contact with the real behavioral world and facing the problems arising there.

Consciousness, now, is just what brings man in contact with the behavioral world. Contrary to what is generally believed, it should not be considered as necessarily related to introspectionism and the study of so-called "states of consciousness."

Perception and the Representation Theory of Consciousness

Some trends in present-day psychology emphasize again the role of cognitive and perceptual functions in behavior. This may be an important step in the direction of a sound concept of behavior. I see a danger, however, in the fact that, frequently, theories of cognition and perception are still working with concepts adopted from introspectionist psychology. The perceptual world is currently conceived as a "representation" *in* the subject of the world outside him. Thus, this representation is a kind of duplicate of the outside world: a duplicate existing *in* the subject. In a recent symposium on cognitive theory and personality functioning contributors repeatedly expressed themselves in such terms of representation. "Cognitive theory," Scheerer (4) argues, "assumes that the organism is in commerce with its environment through the medium of selective representation." Sweet (5) claims that "via the mediate of sensory processes, representations of the environment are obtained." This perceptual representation is considered a "perceptual response present in the organism."

In the field of cognitive and perceptual processes, as in behavioral science in general, it appears to be necessary indeed to return to what have been called the "naturalistic" or primary facts. If not, psychology may lose contact with the specific phenomenon to be explained, i.e., man behaving in his world of concrete life situations. From the psychological or behavioral point of view, nothing corresponds less to reality than to conceive of perception as a representation *in* man of something which exists outside him. Perception is a *direct* contact with the outside world *itself*. No medium, no duplicate, no representation, is interposed between the subject and his world.

This deep-rooted representational theory stems from philosophical and physiological problems. The psychology of sensation, for instance, studies the physiological processes involved in perception. From that standpoint, perceptual responses are in fact "organismic states which the organism carries about in the nervous system and which *represent* the perceived object for me" (5). Philosophy, on the other hand, raises the problem of a *reality* existing beyond the object of cognitive processes. The object perceived is considered only a representation or a duplicate of that unknown reality as it is supposed to exist in itself, independently of the behaving subject. From the behavioral or psychological viewpoint, however, there is no doubt that man perceives the things themselves, that is, the world itself. Perception does not give him a *representation* of the world, but a direct *presence* of the world itself (1, p. 186).

The representational theory is responsible for introducing an erroneous concept of consciousness. Consciousness has been conceived as the enclosed world of inner representations in the subject. In fact, according to the representational theory, there is on the one hand the world of objective, observable things which transcends psychological activity; on the other hand, there is the psychological world of consciousness where the objective world is represented on an enclosed, inner stage. On this enclosed stage the inner representations of outside things are acting in a private and incommunicable play. Thus, representational theory is responsible for introducing a double stage in our world, and even a double world.

CONSCIOUSNESS: EXPOSURE TO THE OUTSIDE WORLD

As a matter of fact, the sciences of behavior are not directly interested in an incommunicable private world of consciousness. Psychology is right in turning to the observable world itself in which man really behaves. But turning to the objective, observable things does not mean denying consciousness. On the contrary, this world of objects, fellow men, and situations precisely *exists for us*, is *present* to us, by the very fact of man's "openness" or exposure to the outside things. This specific exposure constitutes consciousness.

Thus, the naturalistic fact to which behavioral sciences should return is that man, at any moment, *finds himself in a world* of things, fellow men, and meaningful situations. His behavior is nothing other than his commerce or dealing with this world of life situations. In other words, the primary fact to be taken into account in the conception and study of behavior is the fact that an outside world is *present to us*. This presence of an outside world consists in a *direct* contact with things and men outside ourselves. This means that man is not an enclosed structure in himself, but that he is essentially constituted as an "exposure" to an outside world. This concept of exposure and direct presence of the world to us is fundamental in this sense: it underlies the possibility of "meeting" people and things, and thus the possibility of *behavior* as a dealing with situations. Looking at something, speaking to an audience, attending a conference, going away or staying at home, feeling frustrated, etc. mean and imply being in the presence of an outside meaningful world.

This presence of the outside world to us is not to be conceived, of course, as a physical fact (physical neighborhood, for instance). A tree or a room may be situated in the same physical environment as we are, and nevertheless not deal with this environment as with their "world"; they do not manifest that a world is present to them or that they are living in a world. Their responses are limited to physical or chemical reactions to stimuli. Therefore, the presence of a meaningful world to us is a psychological or behavioral fact.

Consciousness in its primary sense

means precisely this exposure of the organism to the world, or in other words the direct presence of a world to us. The very fact, for instance, that an audience is present to me, or exists for me, as a part of my world or situation at the moment I address them, derives primarily from the fact that I am conscious. Consciousness means exposure to the outside world, not enclosedness in the "inner self." Moreover, consciousness or the fact that man finds himself in the presence of an outside world is not to be considered as a "private" affair of inner life. On the contrary, it is an object of direct perception. Seeing a man speaking with other people or driving a car, I perceive the fact that streets, houses, and people are present to this man, that they exist for him, that he is aware of them, that his behaving organism is open to them, i.e., that he has consciousness.

In order to avoid any misunderstanding, it should be stressed that recognition of this primary fact of consciousness as exposure to the outside world does not contradict the theory of perception as a selective and constructive process. The concrete content of the behavioral world present to us is a selective construction of the functioning psychophysiological personality. But no specific content of a perceptual world could ever be built up if there were not this fundamental property and activity of the organism which is called consciousness. In other words, man's exposure to the world is not the result or the total sum of a series of concrete perceptual contents. On the contrary, selective and constructive perception can only be conceived as a continuous activity of specification of the outside world to which man finds himself exposed. This fundamental fact of exposure as such is what is meant by consciousness. Without the fundamental capacity of the organism to be open or exposed to an outside world, perceptual processes would be reduced to internal organismic changes. *Lack* of consciousness, therefore, means enclosedness in the realm of physical and biochemical interactions. Consciousness is psychophysiological activity underlying all psychological or behavioral processes, insofar as they are intrinsically directed toward objects. Even the so-called unconscious motivations and behavior patterns presuppose consciousness in this fundamental sense.

Psychological and physiological sciences, of course, have to investigate the processes involved in the subject's exposure to the world. But the scientific examination of these processes should not destroy the fundamental psychological fact that a world is present to us, or exists for us, and that man's behavior is meaningful dealing with this world. The purpose of scientific psychology is to explain and to account for this fact, not to replace it, for greater convenience, by underlying but different processes. In other words, the underlying processes discovered by scientific investigation should ultimately be related to that unique behavioral phenomenon to be explained. The concept of consciousness does not give a solution to any scientific problem about behavior, but it is necessary for an adequate conception of behavior itself, and for a realistic approach to behavior problems.

Thus, the introduction of the notion of consciousness does not mean going back to introspectionism. On the contrary, it simply means the recognition that the essential characteristic of behavior itself is its being related to a meaningful outside world with which the subject finds himself presented. We also should not forget that scientific research itself is a form of behavior, i.e., a form of dealing with the world we are living in. Communicating scientific results is another form of behavior. Both activities imply what we have called consciousness.

THE VICARIOUS PRESENCE OF THE WORLD IN CONSCIOUSNESS

Until now, we have considered consciousness as putting us in direct contact with the outside world and with the external behavioral facts. The problem of consciousness, however, becomes complicated by the fact that the world can be present to us in different ways. Man is able to find himself in the presence of an object or a situation which is not immediately presented to him from the outside. This is what I call the *vicarious presence* of an object. This possibility of vicarious presence of things to us, via images, concepts, and symbols, is a capital behavioral fact. It enlarges to an enormous extent the world in which man is living and, consequently, the possibilities of behavior. By this vicarious presence man's dealing with situations goes far beyond the limits of responding to actually present things. Man is, for instance, *thinking* about them; i.e., he is dealing with them in a cognitive way. Thinking and imagining are in fact nothing else than other ways of going along with the world and dealing with its situations. This inner, cognitive dealing with situations is intrinsically directed to, or projected toward, an outside world and it is, as a rule, preparatory or compensatory to effective external behavior. In some circumstances—as is the case, for instance, in science—thinking about things actually constitutes the effective form of behavior itself. In fact, man finds himself also in this very particular situation that the world is perceived as a problem to be solved, i.e., to be dealt with in a cognitive way.

MAN'S PRESENCE TO HIMSELF

Man not only finds himself in the presence of an outside world; he may also be presented with himself. His own activities may become for himself the objects of his perception and cognition, as is the case in introspection. In this way man finds himself presented with his own activities, thoughts, and feelings about the world: he is reflecting on himself, on his activities, and on the content of his spontaneous or direct consciousness. This inner presence of man to himself and the reflection upon his own activities may be called *reflective consciousness* or consciousness raised to the second power.

Reflective consciousness is another way in which man escapes "enclosedness" in himself. By the very fact of being able to reflect on himself, man never finds himself completely enclosed in the actual phase of interacting with the environment. Reflecting on what he is and what he does, man "goes out" of himself, transcends the spatial and temporal determinations of the self by judging and evaluating himself and his activity.

At the same time, man's presence to himself enables him to look at himself and at the world from different points of view and to reshape at any time, to a certain extent, his own personality. It enables him to intervene, intentionally, in the course of his own development in order to restructure his relationships with the environment. This possibility seems to be implied in some phases of the therapeutic process (3, p. 422).

Once more, this highest function of consciousness broadens man's world and the self in a considerable way. It frees him from enclosedness in his own subjectivity and enables him to look at things and at himself from an "objective" point of view. This possibility is the basis of such forms of behavior as science and intentional efforts of self-improvement. Thus, consciousness in its different functions frees man from enclosedness in his own organism, in the immediate physical world, and in his own subjectivity. While some degree of exposure to the world seems to be implied in all behavior, the higher functions of consciousness only manifest themselves in specific forms of human behavior (language, culture).

CONSCIOUSNESS AND BEHAVIORAL SCIENCE

As a matter of fact, the vicarious presence of things and man's reflective consciousness of his own activity take place on the level of inner representation. These more complicated functions of consciousness, however, presuppose consciousness in its primary sense of direct presence of an outside world to man.

It is essential for the science of behavior not to separate representational consciousness from direct exposure to the world, as it is important not to separate the representional way of dealing with situations from effective behavior. The representional level of consciousness is not to be studied in itself, as opposed to real behavior, i.e., as a world of "existential" qualities of inner experience. Clinical psychology and projective techniques have shown how closely related these representations and images are with man's real behavior in the world he is living in. Therefore, the introspective data of reflective consciousness and the vicariously present world on the level of representation are to be studied in their functional relationship with behavior as a whole. Even for clinical psychology it is important to base its insights on a behavioral theory in which the represential and operational levels of behavior are functionally related.

On the other hand, the meaningful world itself as it is present to us by direct consciousness cannot be separated from behavior. In fact, the world of meaningful things and situations is a crystallization or condensation of behavioral patterns. A physical object or a part of the surrounding space becomes a desk, a phone, or an office insofar as a whole network of behavioral dealing, and possibilities of dealing, are involved in it (2, p. 447). In this way, consciousness in the objective sense of "the world present to us" is crystallized behavior.

Moreover, man's conscious presence in the world is a *behavioral* presence. This means that, for the human personality, finding itself in the presence of a world implies and entails, necessarily, behavioral dealing with the world. In this sense, consciousness or exposure to the world is already *implicit behavior,* while behavior itself is externalized consciousness or *realized* presence to the world. In any case, instead of opposing consciousness and behavior, one should realize that no human realities are more inseparably united.

CONSCIOUSNESS, PERSONALITY, AND MOTIVATION

Finally, I have to indicate briefly the relationship between consciousness and the concept of personality. Psychologists as a rule conceive of personality as an internal organization of traits, attitudes, aptitudes, and behavioral consistencies. In fact, personality is such an internal organization; but the most essential characteristic of personality is lacking in such a definition. Personality is essentially a structure which goes beyond its internal organization. Its most characteristic feature consists in the fact that an outlook on and an exposure to the world is included in its own constitution. Personality is a way of being and behaving *in* a world that *exists for* the self. The outlook on a world or life situation is an essential constituent of personality itself. In this sense, the fundamental structure of personality is an "ego-world unity." Nothing could more profoundly mutilate the concept of personality than neglecting this aspect of exposure to the world which is an essential component of personality as such. This exposure, as I have said previously, is just what is meant here by consciousness. If, as Lewin says, it is true that the personality governing behavior is a part of the psychological field, it must not be forgotten that the psychological field itself, in the general sense of "living in a world," enters into the make-up of personality.

Personality conceived in this way as an ego-world structure provides us with a sound basis for a theory of human motivation. Motivation theory today tends, so to speak, to situate and localize needs in the organism itself. But from the behavioral point of view, needs are to be conceived as basic patterns of organism-environment or ego-world relationships which actually are required for organism or personality functioning. In fact, personality is made up of different types of such relationships, not only on the level of biochemical organism-environment interaction, but also on the level of cognitive and affective ego-world relationships. While the basic patterns of biochemical interaction—and thus the basic drives in this field—are preformed in the physiological constitution of the organism, the cognitively elaborated ego-world relationships are more personal in origin and nature. Motivations on this level of behavior present themselves as an infinite variety of ego-world relationships in states of tension. Basic needs on this ego level are to be conceived as certain general and significant trends which eventually can be discovered in, and abstracted from, this variety of ways in which the individual tries to insert himself in his world. Such a general trend might be called "need for self-realization," as this "significance" is discovered in, or abstracted from, the innumerable concrete forms of man's behavioral dealing with the world of his fellow men and life situations.

The patterns of ego-world relationships which constitute the infantile personality will be different from the adult ego-world structure. The same significant trends, however, may be discovered in both. In that way, there may be infantile and adult forms of the same basic need, as there are infantile and adult patterns of ego-world relationships. There is no reason why the infantile form should be considered the *basic* one, and the adult pattern the derived or secondary need. Both are rather to be conceived as varieties of a basic type of ego-world relationship which the individual needs in order to maintain and develop the functional organism-environment and ego-world structure which constitute personality at a given stage of its development.

SUMMARY AND CONCLUSIONS

Psychologists who work in the fields of cognition and perception still frequently consider the perceived world as an internal representation of a real external world. This representational theory is strongly rooted in psychology, but it originates in a philosophical speculation which raises the question of a reality underlying the objects of cognition. From the viewpoint of the psychological study of behavior, the perceived world is by no means a representation *inside* of man (or inside consciousness) of some object whose real existence lies outside of him. The perceived world is the real world itself, and perception consists of direct contact with the environment. For this reason the notion of consciousness, as it has been developed within the framework of introspective psychology, is in need of revision. Consciousness does not refer primarily to an internal world made up of representations; it consists of a direct awareness and of an immediate *presence* of the world itself.

Consciousness therefore signifies, before all else, *exposure to the real behavioral world,* and not confinement in an internal world of representation. Consciousness establishes a man as a being who lives in the world, and thus consciousness is an essential aspect of behavior.

The world, however, can also be present to the behaving subject in a *vicarious* way, via representation. Moreover, man can be present to himself by reflective consciousness. These "higher" functions of consciousness also free man from specific forms of enclosedness.

As for the concept of personality, nothing is less in conformity with reality than to conceive of personality as a pure internal structure of traits, aptitudes, etc. Personality signifies a specific modality of existence and of activity which is characterized by the fact of involvement in a situation or in the world. The fundamental personality structure is therefore an ego-world unity, and not an internal organization of an ego. The psychological field itself, more basically, enters into the make-up of personality. Motivation also must be looked at in relation to the fact that man lives in a meaningful world. In fact, basic needs are general types of organism-environment and ego-world relationships which actually are required by the functioning of organism and personality.

The introduction of the notion of consciousness solves no concrete scientific problem regarding behavior, but it is necessary to appeal to the concept in order to formulate and to come to grips with the problems of behavior and of personality in a realistic fashion.

REFERENCES

1. HUSSERL, E. *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie.* Halle: Max Niemeyer, 1913.
2. NUTTIN, J. *Tâche, réussite et échec: théorie de la conduite humaine.* Louvain: Editions Universitaires de Louvain, 1953.
3. ROGERS, C. R. Significant aspects of client-centered therapy. *Amer. Psychologist,* 1946, 1, 415–422.
4. SCHEERER, M. Personality functioning and cognitive psychology. *J. Pers.,* 1953, 22, 1–17.
5. SWEET, A. L. Some problems in the application of cognitive theory to personality functioning. *J. Pers.,* 1953, 22, 41–52.

(Received November 18, 1954)

Psychological Review
Vol. 62, No. 5, 1955

THE DIFFERENCE BETWEEN *Q* AND *R*

JACK BLOCK

Institute of Personality Assessment and Research, University of California, Berkeley

While a good deal has been said by now about the merits and inadequacies of the *Q* approach to personality data, it seems to the present writer that a misunderstanding still persists as to the relationship between *Q* and *R* results. It is still widely believed that the alternative approaches of correlating *variables* and correlating *persons* produce identical, or rather fully convertible, results. A recent paper by Eysenck (4) iterates this point.

It is the purpose of this note to specify the conditions where the equivalence between *Q* and *R* will exist and the conditions where this equivalence will *not* occur. The controversy which has arisen devolves from differing initial assumptions about the nature of personality rather than being an issue in the realm of matrix algebra.

Let us suppose we have by some means finally established or decided upon the *n* basic variables in terms of which personality is to be described or measured. Shall we find at this point that for all possible subgroups of individuals these variables produce equivalent correlation matrices? Or shall we find that the variables interrelate differently in different individuals or "types" of individuals with fundamentally different correlation matrices resulting?

At our present stage of knowledge, our research is guided and our results to some extent limited by the assumption we choose to make as to this ultimate finding. But even now, evidence is available bearing upon this question. Many findings exist wherein the interrelationship of variables differs, in a reliable fashion, as a function of the particular sample investigated. For example, Eysenck (5, p. 205) reports that for Communists, the correlation between direct and indirect aggression is − .94 while for Fascists, the correlation between the same two variables is + .61. For a Neutral group, the correlation is − .64.

If the assumption that variables interrelate differently within different types of individuals is held, therefore, as most consonant with the psychological reality, the way is opened for the possibility of differences between the factors obtained from *R* and *Q* analysis. For *the results of* Q *technique are not convertible into the results of* R *where the subject population providing the data for analysis is heterogeneous with respect to the modes of interaction between variables. Where homogeneity of interaction between variables exists for all individuals within the subject sample, results from* Q *and* R *are transposable.* By heterogeneity of interaction between variables is meant the situation where the pattern of correlations between variables for one subgroup within the total sample differs reliably from the patterns of correlation derived from other subgroups. That is, the variables, X, Y, Z, and so on interrelate differently in one subgroup of subjects than they do in a second subgroup.

Within a sample of subjects homogeneous with respect to such external criteria as age, sex, intelligence, socioeconomic status, and so on, there may still exist a heterogeneity of interaction between variables. *This is particularly true where the variables being interrelated are derived from the personality domain.* Different "styles" of personality organization are known to exist even within samples homogeneous with respect to conventional criteria.

Some illustrations will help make this point clear. The variable "impulsivity" correlated .54 with "suggestibility" in one population (Air Force captains), and − .50 with "suggestibility" in another population (male graduate students). These differences in correlation are highly reliable. If a sample of subjects had been drawn from each population and these samples combined, the correlation between "impulsivity" and "suggestibility" when computed for the combined sample would be zero, or tilted in one direction or the other, as a fortuitous function of the number of individuals included in each subgroup. An *R* factor analysis of each of these subgroups separately or of the combined sample would yield three widely divergent results which could not be reconciled and which in any event would not be revealing of the various psychologically possible relationships present. On the other hand, if persons are correlated in the above illustration (assuming that such correlation is metrically feasible), four types of individuals emerge: those individuals both impulsive and suggestible; those individuals who are neither impulsive nor suggestible; those individuals who, while impulsive, are not suggestible, and finally, those individuals who are not impulsive but are suggestible. Depending upon the continuity or discontinuity of the relationships involved, individuals intermediate to these polar types may be found. These varying personality patterns would seem to have sufficient behavioral implications to warrant their discernment. As indicated above, some or all of these may well be lost by *R* analysis.

Generalizing from the above illustration, it may be shown that where a matrix based upon correlated variables contains zero-order correlations, thus providing little or no possibility of deriving factors, the same data arranged to provide correlations between persons can result in perfect separation of individuals into type factors.[1] While such extreme situations do not exist in nature, approximations to it to a greater or lesser degree do exist—to an extent we have tended to neglect in our investigations in the past.

Another illustration, this time less hypothetical, of the differences in results obtained by *R* and *Q* analyses will further emphasize our basic point. One hundred military officers were assessed over a three-day period and the evaluations of staff psychologists of each subject were codified by means of the *Q*-sort method of description.[2, 3] The staff *Q* descriptions of each officer were correlated person by person and then cluster analyzed by Tryon's method (8).[4] The first four clusters which emerged seemed, in terms of their defining characteristics, to be arranged along an impulse-control dimension. At one end of the underlying dimension were the dashing individuals who cannot or do not modulate their impulses and needs. At the other end of the continuum were the constricted, defensively repressive individuals. Occupying a middle stretch along the dimension were the comfortably conscientious individuals, relaxed and appropriate in their behavior. A group of rigidly conforming men fell between this latter group and the utterly constricted group.

---

[1] Several years ago, Meehl (6) made essentially this same point in an article on the discrimination possibilities afforded by configurally scoring personality items.

[2] This research was supported in part by the United States Air Force under Contract AF 18 (600)-8, monitored by Technical Director, Detachment No. 7 (Officer Education Research Laboratory), Air Force Personnel and Training Research Center, Maxwell Air Force Base, Alabama. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

[3] Personal views or opinions expressed or implied in this publication are not to be construed as necessarily carrying the official sanction of the Department of the Air Force or of the Air Research and Development Command.

[4] The differences between factor and cluster analysis do not affect the argument being advanced here.

Now while the officers defining the two poles of the dimension differed strikingly in their ratings of *rigidity, fluency, breadth of interest,* and so on, they were alike in being described as *sarcastic, guileful,* and *lacking in dependability*. It is thus obvious that a correlation between, for example, the variables "sarcasm" and "rigidity" computed over the total sample of 100 will be affected by the differing ways in which the variables are related for the several groups being improperly combined. In point of fact, this correlation is − .11, clearly insignificant and uninformative. As other data show, the meaning of sarcasm in the diametrical clusters is different. For the impulsive individuals, sarcasm is seen as an additional expression of their unmodulated spontaneity. For the threatened, constricted persons, sarcasm functioned as a means of externalizing inadequacy feelings and projecting hostility. The nature of these relationships would have been lost if the procedure of correlating persons had not been used.

The argument we have here been advancing is by no means new. Stephenson (7), of course, has propagated it for years. Cronbach (3) also has called attention to the possible difference between *R* and *Q* results. Even Cattell (2), who has vigorously argued against *Q,* says that "*Q technique* has its chief use as a classificatory device for finding the subpopulations in a nonhomogeneous population . . ." (2, p. 502).[5]

In calling attention to the difference between the *R* and *Q* approaches, the present writer does not intend to encourage pell-mell abandonment of *R* and total acceptance of *Q*. There are a number of difficulties involved in formulating data suitable for *Q,* while *R* still remains the method of choice for data situations where the sample being evaluated is homogeneous with respect to "divergent" variables. The purpose of this note will fully have been realized if merely it is recognized that *R* and *Q* are not always fully convertible. It is also suggested, along with Beck (1), that at the present stage of personality research, correlating persons deserves a greater emphasis than it has received in the past.

REFERENCES

1. BECK, S. J. The science of personality: nomothetic or idiographic? *Psychol. Rev.,* 1953, **60**, 353–359.
2. CATTELL, R. B. The three basic factor-analytic research designs—their interrelations and derivatives. *Psychol. Bull.,* 1952, **49**, 499–520.
3. CRONBACH, L. J. Correlations between persons as a research tool. In O. H. Mowrer, *Psychotherapy: theory and research.* New York: Ronald, 1953. Pp. 376–388.
4. EYSENCK, H. J. The science of personality: nomothetic! *Psychol. Rev.,* 1954, **61**, 339–342.
5. EYSENCK, H. J. *The psychology of politics.* London: Routledge and Kegan Paul, 1954.
6. MEEHL, P. E. Configural scoring. *J. consult. Psychol.,* 1950, **14**, 165–171.
7. STEPHENSON, W. *The study of behavior.* Chicago: Univer. of Chicago Press, 1953.
8. TRYON, R. C. *Cluster analysis.* Ann Arbor: Edwards, 1939.

(Received December 6, 1954)

---

[5] Incidentally, it is interesting to note that although Eysenck (4) cites this article as supporting the equivalence of *R* and *Q,* Cattell does not definitely commit himself on the question of transposability. Thus, he reports that Burt "*claims* to have demonstrated [the transposability] . . . and *some* statisticians agree with this demonstration" (2, p. 503, italics mine).

Psychological Review
Vol. 62, No. 5, 1955

SOME PROBLEMS IN THE THEORY OF VIGILANCE[1]

JAMES DEESE

The Johns Hopkins University

In recent years there has been a revival of interest in problems that center around the classical notions of attention and set. For the most part, this revival is the result of practical concern with the use of human observers as monitors under conditions of monotony. This rather applied approach, however, has not meant that theoretical problems have been neglected. The British school (2, 3, 10) has been concerned with the theory of monitoring (or, as the British seem to prefer, watch-keeping) almost from the beginning of experimental work on the problem. Theory that has arisen from the current work has tended to ignore the classical problem of the description of subjective states in attention, and has centered around concepts useful in describing the relationship between efficiency in detecting signals and various environmental and organismic conditions.

Largely due to the influence of Mackworth (10), the words "attention" and "set" have been replaced by the term "vigilance." These words—attention, vigilance, and set—by no means refer to the same things, but they have common origins in a class of psychological problems, and it is a fruitless and arbitrary task to attempt to distinguish clearly between them. Mackworth has borrowed the term "vigilance" from Head, and he uses it to mean a high state of readiness to perform adaptive and purposive acts. In operational practice, the term is anchored to such derived measures of performance as probability of response to a given stimulus or inferred thresholds for some stimulus. Vigilance is an excitatory state, and in opposition to it, Mackworth proposes an inhibitory process analogous to the inhibitory processes discussed by Pavlov. Mackworth uses his inhibitory construct in a theoretical description of a general decline in ability to detect signals, a decline that he finds for relatively short periods of visual search. These declines occur when the conditions of search can best be described as monotonous. The major justification for the identification of such a decline in ability to detect with the growth of an active inhibitory state seems to be in the fact that Mackworth has observed an analogy to the phenomenon of disinhibition.

The present paper is concerned with further specifications of a concept of vigilance. The most basic assumption involved in the present theory is that the maintenance of a given level of vigilance in an observer depends to some extent upon stimulus events extrinsic to the observer. Much of this paper will consist of a discussion of alternative hypotheses about the exact nature of the stimulus events sufficient for the control of vigilance. The main emphasis in the present account will be upon excitatory properties, and with respect to the excitatory properties of vigilance there will be little in conflict with Mackworth, perhaps because Mackworth leaves vigilance largely unspecified. The present account will ignore inhibitory properties, except for a brief section intended to show some of the advantages that result from minimizing the importance of an inhibitory construct.

---

[1] The research presented in this paper was performed under Contract No. AF 33(038)-22642 between the Institute for Cooperative Research, The Johns Hopkins University, and the Psychology Branch of the Aero Medical Laboratory, Directorate of Research, Wright Air Development Center.

Before presenting the hypotheses associated with the present conception of vigilance, however, it is necessary to mention briefly the operational framework within which the concepts are to be evaluated. The experimental problem to which the present notions are addressed is one in which observers search for small signals, at near threshold value, the location and exact time of occurrence of which cannot be predicted by the observer. The data (6) reported in the present account come largely from studies of simulated radar watching on a PPI-type cathode-ray tube, though except for special conditions noted below, it is hoped that these studies provide fairly general prototypes for many conditions of search. Certainly the degree of agreement between Mackworth's (10) studies of the clock test and the simulated radar test suggests a fair amount of generality to the experimental situations discussed in the present study.

TWO ALTERNATIVE HYPOTHESES

The first experimental problem to which the present notions were applied was that provided by varying the temporal distribution of signals.[2] It quickly became apparent that there were at least two hypotheses about the mode of action of the temporal distribution of signals upon detectability. Experimental evidence has tended to favor one of these over the other, but since the matter is by no means closed, a brief discussion of each hypothesis is given in the following paragraphs.

The first hypothesis was named the "reinforcement" hypothesis, for the good reason that its fundamental mode of action was built on an analogy to the role of reinforcement in learning. Let us see how the analogy applies to the general problem of search outlined above—that is, to search for small, near-threshold signals which occur at random times and places.

At the outset of a period of search or watch-keeping we may assume that the observer operates at some specified level of efficiency, and, for the moment, we may assume that this level is constant from one period of search to another. We can measure this efficiency in various ways, but for the present let us consider the measure provided by calculating the percentage of signals detected in some period of time. This can be considered as a rough approximation to the momentary probability of detection.

The observer's initial level of detection is determined by a host of factors such as motivation, skill, and the more specific receptor factors. We shall not be concerned with these things except to assume that they vary in a nearly random way from observer to observer and from occasion to occasion. After a probable warm-up period (6), it is assumed that the observer's ability to detect (probability of detection) will go through a simple course of linear decline. Rather than introduce an unnecessary inhibitory construct, we will assume that this represents a dissipation in the initial excitatory state of vigilance. Later it may be fruitful to relate this excitatory state to other determining states, such as motivation, but for the present problem we may leave these relationships unspecified.

It is apparent that we do not always get simple linear declines in probability of detection. This brings us to the critical assumption in the reinforcement hypothesis. The theory assumes that the occasional signals which occur in the observer's field of search determine the future course of his detectability. Specifically, it is assumed that either the *occurrence* or the *detection* of a signal leads to an immediate return of the observer's state of vigilance to its initial level or to a level determined by a change in some other experimental variable.

---

[2] The term "signal" is used throughout this paper to refer to those stimuli that the observer is instructed to detect.

A precise statement of this assumption depends upon whether one indicates that the mere presence of a stimulus or its actual detection (as indicated by a verbal response) changes the level of vigilance. There are reasonable arguments that could be advanced for the view that the signal need not be detected in order to be reinforcing; it has long been suspected that observers maintain too high a criterion for admitting detection when it occurs, and there seems to be some ground for assuming "unconscious," i.e., unrecognized, detection of signals. Fortunately, it turns out that for certain purposes it does not make too much difference which version of the assumption we choose; so for present purposes we may phrase the hypothesis in terms of occurrence rather than detection. In an exact statement of the hypothesis, this assumption leads to more elementary mathematics.

One simple prediction which stems from the reinforcement hypothesis is that signals close together in time will, on the average, have a higher probability of being detected than signals spread over longer time intervals. This hypothesis has been tested in a simple experiment (6), in which observers reported the appearance of small, dim signals presented on a cathode-ray tube in a simulated radar-viewing test. The targets were randomized with respect to locus on the tube. With an important exception, they were also randomized with respect to time. The temporal distribution had the restriction that the average rate of signal appearance was specified and certain maximum and minimum times between signals were imposed. Four different rates were used, a rate of 10 signals per hour, one of 20 per hour, one of 30 per hour, and one of 40 per hour. The percentage of targets detected under each of these rates was measured in half-hour intervals over periods of search three hours in duration.

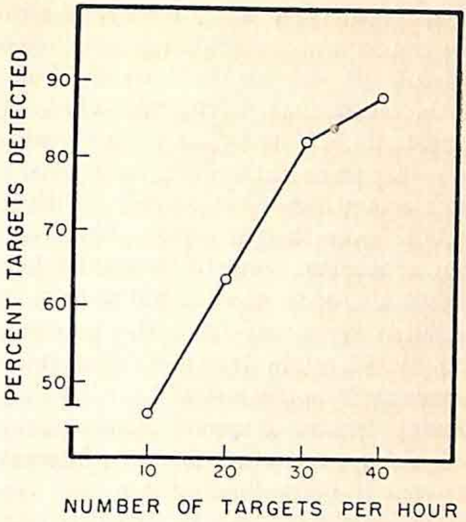


FIG. 1. Per cent signals (targets) detected as a function of the average rate at which targets appeared on the display. The points are averages for four subjects each on four three-hour periods of search.

The results of this experiment can be seen in Fig. 1. It is apparent that the hypothesis seems to be in accord with the facts. Detectability is higher when signals appear more frequently. Several subsequent experiments (6) verified this basic finding; signals appearing at a higher rate seem to be more readily detected. Furthermore, these experiments are in substantial agreement with those performed by Jenkins.[3]

However, another way of looking at the same data provides an additional, independent test of the hypothesis. We can look at detectability as a function of the size of the intersignal interval rather than of the rate of signal appearance. It is by no means necessary that these parameters measure the same thing; indeed, as we will see, they do not. So rather than look at rate, we can see what happens to the probability of detection for a given signal when that signal is separated from the previous one by a certain interval. The reinforcement hypothesis would predict that detectability of a second signal would be greater the smaller the time interval between two signals; that is to say, the reinforcement hypothesis predicts the same thing for both rate and intersignal interval.

---

[3] As reported in an unpublished paper by H. M. Jenkins, read in a Symposium on Vigilance before the American Psychological Association in New York, 1954.

An exact test of this hypothesis demands that we consider the problem of whether it is the occurrence or the detection of a signal that has the postulated action of reinforcement. If we assume that the mere occurrence of a stimulus has a reinforcing effect, we would count intersignal intervals as the time between successive signals presented to the observer. If, on the contrary, we assume that it takes a detected (reported) signal to have the reinforcing effect, we would count the intersignal interval as the time from the last signal detected to a given signal. We have tabulated the data both ways, and while the outcome is not completely unequivocal (6), it certainly is true that the hypothesis is not supported. That is to say, intersignal interval either has *no* effect upon probability of detection (within a given rate) or the effect is in a direction *opposite* to that predicted by the hypothesis. In one experiment there was the suggestion that probability of detection could increase with intersignal interval.

In any case, the experiments from the author's laboratory, as well as analysis of Mackworth's experiments (19), fail to support the reinforcement hypothesis on this point. In view of this fact, the reinforcement hypothesis was abandoned and a new hypothesis explored. Like the reinforcement hypothesis, this hypothesis stems from the notion that a basic excitatory state of vigilance is maintained by conditions of external stimulation. This hypothesis, however, involves a different set of assumptions about how the excitatory state is maintained. These assumptions are embodied in what we may call an expectancy hypothesis, to keep the analogy to learning theory. The hypothesis states that (*a*) the observer's expectancy or prediction about the search task is determined by the actual course of stimulus events during his previous experience with the task, and (*b*) the observer's level of expectancy determines his vigilance level and hence his probability of detection.

Neither of these is a particularly obscure or difficult assumption. Extrapolation from the present behavior of the environment to its future behavior is a common enough activity of man. The accuracy with which observers can discriminate the rate of occurrence of signals does not become a serious question, since the hypothesis is not stated in a way to predict the exact level of detection. The hypothesis simply states that the feedback from the search task determines what the observer expects from further participation in the task in a simple proportional relationship and that his vigilance will vary accordingly.

With this notion of expectation determining output, it is easy to see that one arrives at the same prediction with respect to detectability and average rate of signal appearance as that made by the reinforcement hypothesis. As long as the average frequency of targets remains low, the expectancy of detecting one is low, and when frequency is high, expectancy will be high. With respect to the intersignal interval question, however, the expectancy hypothesis and the reinforcement hypothesis part company. In the form outlined above, the expectancy hypothesis merely predicts that the detection for a given signal should be determined by the average of all signal intervals before the one in question. Thus the probability of detection for any given signal in a series should be the same as that for any other signal, assuming that the distribution of signal intervals is not some nonrandom function of time.

FURTHER DEVELOPMENT OF THE EXPECTANCY HYPOTHESIS

Any further prediction that the expectancy hypothesis might make about the effects of intersignal interval would stem from an extension of the first assumption of the hypothesis. The first assumption of the expectancy hypothesis states that the observer's prediction about the future course of stimulus events in the field of search is determined by the actual course of stimulus events in his past experience. Thus the assumption demands a linear relationship between the average of some selected aspect of past events and vigilance with respect to events at any given moment. If the task is invariant with time, vigilance will change little after the average level of expectancy has been reached. It is probable, however, that such an assumption demands that the observer's activities be very much more simple than they probably actually are in search experiments. In addition to an over-all level of expectancy determined by a nonweighted composite of past events, it is reasonable to suppose that the observer's expectancy fluctuates over short ranges of time in a predictable manner.

Let us assume, for example, that for any given signal the observer's expectancy is exactly determined by the mean intersignal interval. Let us further assume that the observer's ability to estimate time since the last target is free of constant error and shows relatively low variable error. If these assumptions are correct, expectancy should be low immediately after a signal, should increase as the mean intersignal interval is approached, and finally should become quite high as the intersignal interval grows beyond the mean. The mean probability of detection averaged over these periodic fluctuations would correspond to the average intertarget interval. This notion implies that if we look at probability of detection as a function of intersignal interval, we ought to obtain something quite different from the prediction made from the reinforcement hypothesis. Unfortunately, the data thus far available are equivocal, and the equivocality does not only depend upon the confusion caused by the alternate assumptions of occurrence or detection affecting vigilance. There is some evidence from Mackworth (10) and from experiments by the present author (6) indicating that probability of detection increases with increasing intervals between *presentation* of signals; unfortunately, however, both Mackworth's data and the data from the author's experiment are confounded, for purposes of this analysis. One experiment in which spatial location and frequency of intersignal intervals were not confounded with size or intersignal interval obtained negative results (6), both on analysis by stimulus presentation and interval following stimulus detection. A more thorough experimental exploration of the problem is still in order, and it seems safest to regard the issue as unsettled.

Perhaps it should be pointed out that in regard to the question of the temporal distribution of signals, the differences between the two hypotheses we have been discussing are mainly concerned with one rather elementary point. The question is whether the probability of detection for any given signal is determined exclusively by the temporal location of the signal immediately preceding the signal in question or whether it is determined by a large and rather indeterminate number of signals preceding the signal in question. The assumed activities on the part of the observer differ principally in the nature of reaction to signal feedback. The failure of the reinforcement hypothesis suggests that it is not fruitful to regard the observer as a passive instrument that remembers only the time since the last signal. Rather, it seems more fruitful to regard the observer as a detecting instrument that is continuously performing a kind of averaging of previous input in order to extrapolate the results to future behavior of the search field. This process can be considered to be directly analogous to the important notion of adaptation level, hypothesized by Helson and his associates (e.g., 7).

Let us briefly consider some problems other than temporal distribution to which the expectancy hypothesis might apply. For example, there is the problem of the spatial distribution of signals. For an experimental hypothesis, we might suppose that the probability of detection for different portions of the field of search could be altered by systematically varying the spatial distribution of signals in the field. In point of fact, it was rather a disappointment to find that no such effect could be experimentally induced (6). It could be argued, however, that the experiment conducted by the present author was inadequate; the entire field of search covered only about 30 degrees of visual angle, and the absolute frequency of signals was very high (60/hour), thus minimizing the possibility of differences attributable to spatial distribution of previous signals. Finally, it is possible that the spatial distributions were not extreme enough. Despite all of these arguments as to the inadequacy of the experiment, it seems likely, for radar viewing, that such an effect would be extremely inconsequential in magnitude, if it occurred at all.

The reason for this tentative conclusion stems from an earlier experiment (4) from the author's laboratory. In this experiment detectability thresholds (in decibels of attenuation from a fixed signal voltage) were measured under a variety of conditions in which observers were (*a*) instructed to search only a portion of the scope, and (*b*) forced to search only a portion of the scope by having the remainder of the scope masked over. Instructing observers to search only a portion of the scope had no effect upon detectability, but one could improve detectability to the extent of about 2 db. by masking out a portion of the scope (if the masking is done by annuli rather than sectors). From these results it seems unlikely that detectability for PPI-type displays would be very sensitive to changes in spatial distribution of signals, unless the observers were forced not to attend to part of the search field. This fact is probably an inherent property of the viewing of PPI displays, and most likely reflects the effect of a moving sweep and luminance gradient. The result of these factors is that almost inevitably the observer distributes his viewing almost equally over the entire scope. Thus, while some effect of spatial distribution may occur for some visual search tasks, it probably does not occur for all of them.

It is apparent that many problems could arise out of the present statement of the expectancy hypothesis of search behavior. In this brief paper, however, we will look at only one further problem, one that is fundamental to the hypothesis. This problem is provided by the fact that the hypothesis is stated in such a way that the observer's extrapolation of the future behavior of the search field is separated from his vigilance for signals in the field. Such an assumption is necessary if one wishes to avoid the rather artificial implication that vigilance is determined directly by task feedback. It is clear that there are states characteristic of the individual that would not affect his ability to estimate what he *sees the task to be* but which would affect his motivation to participate actively in the task. We may characterize such states as individual differences in vigilance. Having so characterized them, it is our task to discover measures of behavior which predict levels of vigilance expected in some individual on a search task. That such a task is psychometrically feasible is indicated by Jenkins'[4] finding that the within-individual–between-session correlations of probability of detection are high.

The most direct approach to this problem is to look for other measures of behavior in the search situation itself. Among those that have been measured are frequency of false reports (incidence of reporting a signal when none is presented) and reaction time. There are ample data to show that while frequency of false reports varies from one individual to another, it does not correlate with probability of detection (6). On the other hand, Jenkins' analysis of his own data suggests that the reaction time to a pilot signal and probability of detection of the scored signals correlate to a significant degree. This is particularly encouraging, since one would clearly expect vigilance to determine reaction time to a related but unscored signal, and it is entirely reasonable that such a correlation could be independent of the observer's estimate of the future behavior of the task.

It is perhaps a little puzzling that frequency of false report turns out to have no significant correlations (except with rate of signal appearance), but it is at least satisfying (if somewhat *post hoc*) to suggest that expectancy based upon experience swamps any individual differences in tendency to respond to signals. This tentative notion receives some support from the fact that, in the experiments by the present author (6), those individuals who showed a very high false report rate (by verbal report at the close of the experiment) had no higher subjective estimates of the rate of signal appearance than those observers who showed no false reports, or only one or two, during the course of the experiment.

IS AN INHIBITORY CONSTRUCT NECESSARY?

Let us take a brief look at the role of an inhibitory construct in a theory of monitoring or watch-keeping. A general notion of "fatigue" has of course frequently turned up in research on visual tasks. The difficulty of finding a suitable index of fatigue in visual work, however, is ample evidence that the concept, as widely used, is so vague and nonoperational as to be virtually useless. Indeed, Bartley and Chute (1), in their monograph on the subject, showed that they were convinced that the only meaningful referent for the word "fatigue" was to its subjective properties—i.e., to what observers said they felt like.

In the work on visual monitoring, however, a fatigue-like concept has been used in a much more specific and operational context. In Mackworth's studies of visual search he was able to show, under controlled circumstances, that there was a more or less monotonic decline in detectability with time at search over a period of two hours. Various experimental variables introduced by Mackworth convinced him that the decline in performance could be attributed to an active inhibitory process, analogous to the inhibitory processes invented by Pavlov.

---

[4] See footnote 3.

Mackworth's notions were elaborated by Broadbent (3) in a paper on classical conditioning and human watchkeeping, and Broadbent pushed the analogy to Pavlov's inhibitory constructs even further. Both Broadbent and Mackworth assume that a continuous decline in detectability is the rule for visual search under monotonous conditions. However, a difficulty with any simple application of the inhibitory construct seems to be that sometimes detectability rises instead of falls with continuous search (6). Sometimes the curves of detectability as a function of time at search fluctuate in a highly irregular manner (6).

It is easy enough to attribute such fluctuations to recovery from inhibition, in the form, perhaps, of disinhibition. But is more difficult to point out just why disinhibition occurs where it does, or, to be more accurate, it is difficult to predict when and where in a given search task disinhibition is apt to occur, unless one introduces special stimulating circumstances to produce it. Thus we must decide, at the least, that the circumstances determining the development of inhibition are very complicated, or that it is easily obscured even in very simple, well-controlled search tasks.

The case against a special inhibitory construct is further bolstered by a consideration of the nature of the disinhibitory stimuli used by Mackworth. In the strict Pavlovian usage, a disinhibiting stimulus is a neutral one; it is hard to know what "neutral" means in the context of visual search, but it is clear that Mackworth's stimuli were meaningful and perhaps motivating to his subjects.

It is also apparent that in visual search an end-spurt can occur if the observer knows something of the duration of the search period, and there is also the suggestion that a warm-up effect occurs, even when the observers are fairly well practiced at the task (6). While there is no compelling reason not to describe all of these effects in terms of the dissipation of inhibition, there is likewise no good reason for doing so. It seems merely cumbersome to propose a state of vigilance, and then to assume that apparent declines in vigilance are due to the active inhibition of the positive process by an additional, negative process.

There is a widespread belief—so widespread, as a matter of fact, that it has almost reached the proportions of folklore—that continuous, monotonous visual search (as in driving on a modern highway) produces a condition akin to hypnotic sleep (whatever that may mean). At any rate it seems clear, because it can be established in the laboratory, that observers may lose vigilance beyond the point of voluntary recovery. That is to say, no matter how hard they try, they cannot increase their attentiveness to the signals being searched for, whereas an extraneous event may restore their ability to observe and detect signals in the visual field.

Despite the fact that at first glance such effects seem most attractively described in terms of inhibition, a more serious glance at these facts, together with certain auxiliary information, suggests that they are more fruitfully conceived of as the absence of some normal, excitatory state. If we turn to the neurophysiological observations on sleep, we see some interesting possibilities that perhaps have a bearing on the problem of vigilance in search. We know, for example, that the waking center in the hypothalamus, in part at least, must be maintained at a high excitatory level by continued sensory input. Thus the reduction in the sensory input characteristic of monotonous search may reduce the excitability of this or analogous centers in the control of the activity of the organism.

A glance at Kleitman's (8) double theory of sleep shows some interesting comparisons. Kleitman makes a distinction between a wakefulness of necessity and a wakefulness of choice. Wakefulness of necessity is characterized by the polyphasic sleep patterns of many animals and human infants. Whenever such an organism is deprived of sensory input, or the sensory input declines, the activity of the waking center is reduced and the organism sleeps. But to a certain extent the cerebral cortex can control the pattern of wakefulness and sleep; hence the matter of wakefulness of choice. It seems entirely reasonable that the wakefulness of choice depends not only upon the autonomous activity of the cortex, but upon the level of sensory input as well.

This interpretation suggests a good deal more for the theory of vigilance than a mere negation of an inhibitory construct. It suggests the importance of varied sensory input in the voluntary control of vigilance. It is quite possible, for example, that restricting the proprioceptive and tactual input by forcing the observer to remain seated in a rigid position contributes to the decline of vigilance and to an inability of the observer to recover vigilance voluntarily. This is very close to the interpretation given by McFarland, Holway, and Hurvich (9) to their finding that DL's for luminance increase in size as one continues to take measurements; and that, if one allows the observer to stand up and stretch, the DL's recover.

The present view is, therefore, that the problem of maintaining vigilance under conditions of monotonous search can better be viewed as a problem in maintenance of a background sensory input than as the problem of dissipating an accumulation of central inhibition. There are several consequences of this view that are susceptible to experimental test. It suggests that we ought to look for individual differences between observers in measures that reflect an indication of tension level in the individual. Hence, there is the strong suggestion that we look at the interaction between indices of muscle tension, detectability, and signal rate. Furthermore, since there are reasons for suspecting that various personality measures reflect tension levels in the individual, there are reasons for looking at the correlations between personality scores and detectability. The highly promising work of Davis (5), from the Cambridge laboratories, suggests that this is an especially fruitful approach. At any rate, all these problems stem from a notion of vigilance as maintained by excitatory input rather than by inhibition. And, until we have explored these possibilities, it is probably not the best strategy to concentrate exclusively on the properties of an inhibitory construct to account for decline of visual search.

REFERENCES

1. BARTLEY, S. H., & CHUTE, E. *Fatigue and impairment in man.* New York: McGraw-Hill, 1947.
2. BERLYNE, D. E. Attention, perception and behavior theory. *Psychol. Rev.*, 1951, **58**, 137–146.
3. BROADBENT, D. E. Classical conditioning and human watch-keeping. *Psychol. Rev.*, 1953, **60**, 331–339.
4. BUCKLEY, BARBARA B., HANES, R. M., & DEESE, J. Search area and target detectability on a PPI cathode ray tube. *USAF Tech. Report, WADC–TR–503,* 1953.

5. DAVIS, R. Neurotic predisposition and the disorganization observed in experiments with the Cambridge Cockpit. *J. Neurology, Neurosurgery & Psychiatr.*, 1946, **9**, 119–124.

6. DEESE, J., & ORMOND, ELIZABETH. Studies of detectability during continuous visual search. *WADC Tech. Report, WADC–TR–53–8,* 1953.

7. HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 297–313.

8. KLEITMAN, N. *Sleep and wakefulness.* Chicago: Univer. of Chicago Press, 1939.

9. MCFARLAND, R. A., HOLWAY, A. N., & HURVICH, L. M. *Studies of visual fatigue.* Cambridge: Harvard Graduate School of Business Administration, 1942.

10. MACKWORTH, N. H. Researches on the measurement of human performance. *Med. Res. Council, Special Report No. 268,* London: His Majesty's Stationery Office, 1950.

(Received January 7, 1955)

Psychological Review
Vol. 62, No. 5, 1955

STATISTICAL THEORY OF DISTRIBUTIONAL PHENOMENA IN LEARNING

W. K. ESTES[1]

Indiana University

One aspect of distribution of practice that recent reviewers (**14**, p. 189; **24**; **25**) seem agreed upon is its persistent refractoriness to any general theoretical interpretation. If *general* is taken to mean *comprehensive,* then I can see no ground for optimism on this point. In all likelihood distributional phenomena are influenced by an extensive set of variables which vary in relative importance from one situation to another. There is no reason, however, why this complexity should be expected to bar progress toward theories which are limited in the sense of dealing with effects of independent variables taken singly or in limited combinations, yet general in the sense of applying to a wide range of situations.

The approach to problems of distribution that will be developed here continues a theoretical program outlined in a recent paper (**6**). The over-all strategy is to start by analyzing in detail the variables of most ubiquitous influence in order to see how much they account for, leaving for later consideration local variables which are important only in restricted types of experimentation.

As in our previous investigation of spontaneous behavioral changes (**6**), preliminary analysis of the distributional problem discloses two independent variables associated with intertrial intervals that probably modify the course of learning under all experimental arrangements. These variables are stimulus fluctuation and interpolated learning. In the preceding study, the former was incorporated into a mathematical model in order to permit derivation of its testable consequences. Since experimental arrangements are essentially the same in studies of spacing as in studies of spontaneous changes, we need only make further application of the same model in order to investigate the role of stimulus fluctuation as a determiner of spacing effects in learning.

It will be convenient to consider separately the two principal aspects of learning distinguished in contemporary theories: (*a*) changes in resistance to extinction or resistance to forgetting, and (*b*) changes in momentary probability, rate, or latency of the response. This distinction will be recognized as corresponding in essentials to that between habit strength and momentary reaction potential in Hull's system (**10**), or that between reflex reserve and reflex strength in Skinner's (**21**). For brevity I shall use the terms *habit strength* and *response strength* to distinguish the two concepts.

At a common-sense level, we should expect that stimulus fluctuation would affect the two types of learning measure differently. To the extent that fluctuation occurs, the make-up of a stimulating situation will vary from trial to trial or from period to period. Therefore, insofar as the effects of this one variable are concerned, rate of learning in terms of change in response strength should be increased by reducing the intervals between trials or periods. On the other hand, resistance to extinction or forgetting should be expected to depend on having the response conditioned to as many as possible of the stimuli which might occur during extinction or during a test for retention; therefore, rate of change in habit strength should be directly related to intertrial interval. These notions will be worked out more rigorously in the following pages.

---

[1] This paper was prepared during the author's tenure as a faculty research fellow of the Social Science Research Council.

The concepts and assumptions that will be needed for this task have all been presented in a previous paper (6). Only those that are most critical for our present purposes will be reviewed briefly at this point. In order to permit self-selection diets for readers of varying mathematical appetites, I shall again follow the procedure of discussing the model and its applications first at a verbal level and then, in a separate section, in mathematical terms.

CONCEPTS AND ASSUMPTIONS

It is assumed that the stimulus variables associated with any environmental situation undergo constant, random fluctuation. Consequently, of the total population of stimuli available in a given situation over the course of an entire experiment, only a randomly selected subpopulation will be available during any one experimental period. On any one trial, the organism samples the available subpopulation; the probability that any available stimulus element will be sampled on a trial is the parameter θ which has been prominent in other developments of statistical learning theory (7, 19). We shall represent the total stimulus population associated with any experiment by a mathematical set S_*, the subpopulation available on any one period by a set S, and the subpopulation of temporarily unavailable elements by S'.

There is, then, a continuous, random interchange of elements between S and S'. During a given period, all the elements in the available set S may be conditioned or extinguished while those in S' are unaffected. Since probability of a given response is assumed equal to the proportion of conditioned elements in S, it is possible for response probability to go to unity or zero during a period of conditioning or extinction even though there remain in S' many elements, unconditioned or conditioned, respectively, which may reappear in S during some later period and alter the response probability.

INTERPRETATION OF DISTRIBUTIONAL PHENOMENA

Rate of change in habit strength as a function of spacing. According to our assumptions, there will be available for conditioning or extinction during any one period [2] only a limited portion of the total population S of stimulus elements which may appear in the given experimental situation at any time. Consequently, complete extinction of a response can be achieved only if the series of extinction periods is distributed over a long enough interval so that in the course of random fluctuation every previously conditioned element in the population appears in the available set S during at least one period. And in general, the greater the number of conditioned elements present in both parts of the population S_* at the onset of extinction, the longer it will take, in terms of time, trials, or periods, to reach a given criterion of extinction. We have also seen in an earlier paper (6) that, according to the present theory, resistance to forgetting will be directly related to the total density of conditioned elements in S_*.

Since both of the observable symptoms of habit strength vary directly with total density of conditioned elements, we need only consider the relationship between the latter variable and conditions of spacing. The predicted effects of varying the length of the rest interval between successive conditioning periods follow directly from the sampling model. Rate of increase in the proportion $F(m)$ of conditioned elements in S_* will be maximal if the interval is long enough so that the available situation S during each successive period represents a new random sample from the total population. As the interval is reduced, there is less opportunity for elements that are in S, and thus become conditioned, during one period to escape and be replaced by unconditioned ones before the next period, and consequently rate of growth of $F(m)$ is retarded. These considerations give rise to the family of curves illustrated in the lower panel of Fig. 1, which has interperiod interval as a parameter. Following the first period, steepness of the curve is a uniformly increasing function of the interval. The negative acceleration of these curves is due, of course, to the fact that there will be fewer unconditioned elements remaining to be sampled after each successive period. Analogous considerations hold for changes in $F(m)$ during extinction so that, other things equal, the "resistance to reconditioning" will be an increasing function of the interval between successive extinction periods.

---

[2] To avoid circumlocution, the term *period* will be used here to refer to any of the following: (*a*) a period during which conditioning or extinction is carried to a criterion, (*b*) a period of fixed duration, (*c*) a block of trials, or (*d*) even in some circumstances a single trial.

Rate of change in response strength as a function of spacing. Changes in response strength during learning may be followed empirically by obtaining measures of response probability, rate, or latency at the beginning of each successive period. For simplicity, we will consider only probability. It can be shown (5) that all arguments that hold for probability carry over directly to the other measures.

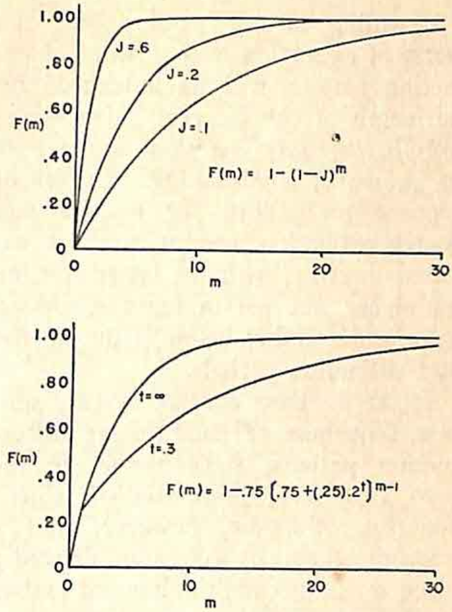


FIG. 1. Proportion of conditioned elements, $F(m)$, in the total population as a function of conditioning periods, m. In the upper panel the parameter is J, the relative size of the available set S during any period; in the lower panel the parameter is t, duration of the interperiod interval.

Rate of change in response probability during learning will be less simply related to the interperiod interval than was the rate of change in habit strength. If we compare probability curves for a relatively massed and a relatively spaced group, the former will certainly be expected to show the steeper rise at the outset of conditioning, for between the first and second periods fewer of the newly conditioned elements will escape from S if the interval is short. On the other hand, the gain per period in total number of conditioned elements is greater for the spaced group. From this fact we might guess that as the series goes on, the theoretical curve for the spaced group will overtake and then surpass the curve for the massed group. As a matter of fact this surmise can be proved to be correct, as we shall demonstrate in a later section. The principal characteristics of the family of curves having interperiod interval as a parameter are illustrated in Fig. 2.

According to our assumptions, the course of extinction over a series of extinction periods will be influenced by the length of the interval between periods in precisely the same manner as the course of conditioning. Curves of response probability for massed and spaced extinction groups will be expected to cross, with the latter starting out higher but overtaking the massed curve and running below it during the later extinction periods.

So far we have considered only uniform sequences of conditioning or extinction periods, since these are the cases that have been studied experimentally. Suppose, however, that a group of organisms were administered a mixed series of conditioning and extinction periods. Predictions can be generated from the present theory concerning the course of learning over any specified series. An interesting special case arises if we modify the usual partial reinforcement procedure by taking a block of trials or an experimental period rather than a single trial as the unit relative to which reinforcement is randomized. A theoretical account of this case is given in the following section and illustrated in Fig. 2. An experiment which provides a preliminary test of the theoretical predictions has recently been reported (12).

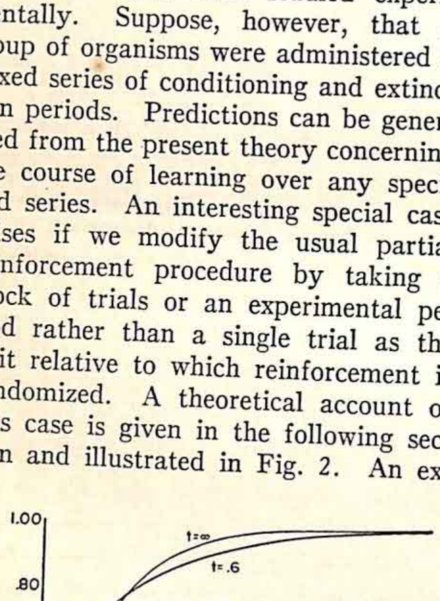


FIG. 2. Response probability in the temporary situation S as a function of number of conditioning or extinction periods, m. The two upper curves represent a sequence of conditioning periods, the two lower curves a random sequence of conditioning and extinction periods with π equal to .5. In each case the parameter is t, the interperiod interval.

MATHEMATICAL DEVELOPMENT OF DISTRIBUTIONAL THEORY

Now it is desirable to show more rigorously how the assumption of random stimulus fluctuation leads to detailed quantitative predictions concerning spacing phenomena. A mathematical model incorporating the fluctuation concept has been presented in a preceding paper (6). Here we shall review briefly a few important terms and relations, and then proceed to apply the same model to problems of spacing.

As before it will facilitate exposition to restrict the formal treatment to highly simplified conditions. We will assume that the behaviors available to the organism in a given situation are categorized into two mutually exclusive and exhaustive classes, A and $\bar{A}$, the former being the response recorded; that each stimulus element in S_* is conditioned either to A or to $\bar{A}$; that probability of response A at any time during an experimental period is equal to the proportion of elements in S conditioned to A; and that all elements in S^* have equal probabilities of appearing in S.

Let t represent duration of the interperiod interval; N and N' the mean numbers of elements in S and S', respectively; C the probability that an element in S' during period $m-1$ appears in S during period m; and B the probability that an element in S during period $m-1$ is still there during period m. Two important relationships derived in the preceding paper (6) are

$$C = J(1 - a^t) \qquad [1]$$

and

$$B = J + (1 - J)a^t, \qquad [2]$$

where J is the ratio of N to $N + N'$, and a is a slope constant depending on rate of stimulus fluctuation.

Now we are ready to derive predictions concerning the course of learning as a function of trial spacing.

Rate of change in habit strength as a function of trial spacing. We require the probability $F(m)$ that an element is conditioned to response A after m learning periods when the interperiod interval is t and all elements present in S during any given period become conditioned to A. The latter assumption implies that the proportion J of the unconditioned elements in S_* will become conditioned during the first period and the proportion C on each subsequent period, yielding the relations

$$\begin{aligned} F(1) &= (1-J)F(0)+J \\ F(2) &= F(1)+C[1-F(1)] \\ &= (1-J)F(0)+J+C \\ &\qquad \times[1-(1-J)F(0)-J] \\ &= 1-(1-J)[1-F(0)] \\ &\qquad \times(1-C) \end{aligned}$$

and so on. It can be proved by mathematical induction that in general

$$F(m) = 1-(1-J) \times[1-F(0)](1-C)^{m-1}.$$

Since from equation 1,

$$C = J(1-a^t)$$

we have finally

$$F(m) = 1-(1-J)[1-F(0)] \times(1-J+Ja^t)^{m-1} \quad [3]$$

and it can be seen that $F(m)$ is an increasing function of the interperiod interval t for all m. This function is illustrated in the lower panel of Fig. 1 with t as a parameter. Under the conditions assumed here, varying the size of the training unit, i.e., the duration of a conditioning period or the number of trials per daily block, will simply result in modifying the value of J, yielding the family of curves illustrated in the upper panel of Fig. 1. Analogous considerations hold for changes in $F(m)$ during extinction.

Comparison of the upper and lower panels of Fig. 1 might lead one to surmise that, beyond a moderate rest interval, the advantage of distribution of practice will come chiefly from dividing a given number of training trials into a larger number of spaced blocks. As a matter of fact this prediction does follow from the theory, not only for the highly restrictive conditions assumed in the derivations of this paper, but quite generally. For any situation in which the learning curve within a daily period is negatively accelerated, the total number of trials needed to attain a criterion of learning will increase with the number of trials per daily period, the function approaching linearity as the number of trials per period becomes large. This relationship is attributable to the fact that fewer elements are conditioned on the second trial of a daily period than would have been conditioned if this trial had been postponed until the next period, and a similar argument holds for each succeeding trial. Once the daily period is made large enough so that nearly all of the available elements are conditioned to the response in question, further increase in the number of trials per period will simply yield a proportionate increase in the total number of trials to a criterion.

Rate of change in response probability as a function of trial spacing. Let us represent by $p(t,m)$ the probability of response A in S at time t following the mth of a series of learning periods, the periods being uniformly spaced with interperiod interval equal to t. First we consider a series of acquisition periods with all elements present in S during a period becoming conditioned to response A. We can write immediately:

$$p(t,m) = \frac{1}{N}[NB+\{(N+N')F(m)-N\}C]$$

where B, C, and $F(m)$ are as defined above. The first term inside the brackets is the number of conditioned elements in S at time t that are held over from the preceding experimental period, while the second term is the number of conditioned elements that were outside S at the end of the preceding period but drifted into S by time t. Simplifying by means of the relation $J = N/N + N'$, and expanding C and B by means of equations 1 and 2, we arrive at the function

$$\begin{aligned} p(t,m) &= B + \left(\frac{F(m)}{J} - 1\right)C \\ &= 1 - (1 - J)[1 - F(0)](1 - a^t) \\ &\qquad \times (1 - J + Ja^t)^{m-1} \qquad [4] \end{aligned}$$

which for all its forbidding appearance has only three free parameters and is relatively simple in form, as illustrated in Fig. 2. It will readily be seen that when m equals 1, response probability will be inversely related to the interperiod interval t. For sufficiently large m the derivative of $p(t,m)$ with respect to t will be positive and therefore response probability will be directly related to t. Therefore, as shown in Fig. 2, any pair of otherwise comparable massed and spaced curves will cross exactly once.

The corresponding function for a series of extinction periods during which all elements present in S are unconditioned from A, i.e., conditioned to an opposed response $\bar{A}$, is obtainable in precisely the same way and turns out to be

$$\begin{aligned} p(t,m) &= F(0)(1 - J) \\ &\qquad \times (1 - a^t)(1 - J + Ja^t)^{m-1} \qquad [5] \end{aligned}$$

Random reinforcement. Finally we consider a series of periods run under the same conditions assumed in the derivation of equations 4 and 5 except that acquisition and extinction periods appear in a random order. It can be verified by substitution from equation 4 that if period $m + 1$ is an acquisition period then

$$p(t,m+1) = (1 - C)p(t,m) + C$$

whereas if it is an extinction period, it follows from equation 5 that

$$p(t,m+1) = (1 - C)p(t,m).$$

Again C is being used as an abbreviation for $J(1 - a^t)$. If acquisition and extinction periods occur with relative frequency π and $1 - \pi$ respectively, the expected response probability after period $m + 1$ will be the average of the two expressions preceding:

$$\begin{aligned} E[p(t,m+1)] &= \pi[(1 - C) \\ &\quad \times p(t,m) + C] + (1 - \pi) \\ &\quad \times (1 - C)p(t,m) \\ &= (1 - C)p(t,m) \\ &\quad + \pi C. \qquad [6] \end{aligned}$$

Solving this difference equation by standard methods we obtain

$$p(t,m) = \pi - K(1 - C)^m$$

where K is an arbitrary constant, and finally

$$\begin{aligned} p(t,m) &= \pi - (1 - J)[\pi - F(0)] \\ &\qquad \times (1 - a^t)(1 - J + Ja^t)^{m-1} \qquad [7] \end{aligned}$$

when K has been eliminated by setting the expression for $p(t,1)$ obtained from equation 7 equal to the mean value of $p(t,1)$ computed from equations 4 and 5. Since the quantity $(1 - J + Ja^t)$ is a fraction between zero and one, the second term on the right side of equation 7 tends to zero as m increases, and response probability approaches π as an asymptote regardless of the value of J or the duration of the interperiod interval. The dashed curves of Fig. 2 illustrate the effect of varying intertrial interval in a series with $\pi = .5$.

It should be noted that the expressions for response probability derived in this paper refer to expected values for individual organisms or for groups of organisms with like values of the parameters: J and a. However, it is not likely that the errors of approximation involved in applying the derived functions to averaged data from groups of experimental subjects will be large relative to the errors of measurement involved in ordinary learning experiments.

EMPIRICAL RELEVANCE AND ADEQUACY

What I have tried to accomplish in this paper and the one preceding (6) is to work out in systematic fashion the behavioral effects of stimulus fluctuation which is strictly random with respect to experimental variables. To the extent that distributional phenomena prove to be attributable to this source of variation, the burden of explanation upon such constructs as inhibitory potential will be commensurately reduced. To be sure, the conclusions we have reached are not independent of the specific techniques of conceptualization and derivation used, but it seems doubtful that the main features will prove very sensitive to the choice of mathematical techniques.[3]

In applying the theory to learning experiments, we are led to fractionate the empirical material into three main groups: (*a*) experiments involving continuous work or very short rest intervals; (*b*) experiments involving intermediate intervals, ranging from perhaps a minute up to a day or two; and (*c*) experiments involving long intervals. It is in the intermediate range that stimulus fluctuation theory should be expected to play a major role in accounting for experimental phenomena. At very long intervals, or under certain special conditions, such as those of retroactive inhibition experiments, effects of interpolated learning become important; space does not permit discussion of these effects in the present paper. At very short intervals, variables come into play that have not been taken account of in the statistical theory, e.g., response-produced stimulation, physiological changes in effectors; and even if the predicted effects of stimulus fluctuation are present they will be masked by effects of other variables.

Two principal methods are available for assaying the empirical adequacy of the theory. The first is to carry out new experiments especially designed to satisfy the simplifying assumptions made in the theoretical derivations. The second is to review the experimental literature and attempt to provide a unified theoretical account of the already established facts concerning the role of spacing at intermediate intervals.

Only the former method can be entirely free of bias, and consequently publication of the theory has been delayed until several of the indicated experiments could be carried out. One of these was a study of spontaneous recovery of a bar-pressing response in relation to degree of extinction and amount and spacing of reinforcement (8); another was a study of successive acquisitions and extinctions of a jumping habit in relation to schedule of reinforcement (13); a third was a study of successive reversals of a left-right discrimination with varying number of trials per daily period (11). Detailed theoretical interpretations of the data of these experiments have been reported elsewhere. It is to researches of this kind that we must look for sharper delimitation of the area of application of the theory and for indications as to desirable lines of revision or supplementation.

---

[3] As a matter of fact, in an earlier version of the present model, the assumptions were expressed in the form of differential equations and the theorems were derived by means of the calculus. Subsequently I found it desirable for reasons of mathematical rigor and elegance to rework the theory utilizing the method of linear difference equations which has been introduced into learning theory by Bush and Mosteller (2), but none of the testable consequences of the theory were altered. The earlier formulation was presented to a Colloquium at the University of Illinois in May, 1950, and to the SSRC Summer Seminar on Mathematics Models for Behavior in 1951, and I have benefited greatly from the discussion and criticism contributed on those occasions.

A comprehensive review of the literature on distributional phenomena cannot be attempted here, but it will be noted that a variety of rather well-established empirical relationships can be given a uniform and coherent interpretation in terms of theoretical relationships derived in this paper. It may be helpful to bring together at this point a brief summary of these relationships together with appropriate references to the experimental literature.

a. Resistance to extinction is directly related to number and spacing of conditioning periods (**8**, **17**, **18**).

b. Retention is increased by spacing of learning periods, the advantage of spacing increasing with the retention interval (**14**, pp. 156–158; **16**, p. 508).

c. Retention is increased by overlearning (**9**; **14**, pp. 377–380).

d. Learning tends to be faster at beginning of series with massed practice, later in series with spaced practice: Conditioning (**1**, **22**); Extinction (**4**, **20**, **23**).

e. Number of trials to asymptote of learning is directly related to number of trials per daily period (**3**, pp. 155–156; **15**, p. 320).

f. Retention is directly related to the proportion of common elements in learning and recall situations (**9**; **14**, pp. 449–450; **19**).

The rather surprising prediction that asymptote of response probability under random reinforcement should be independent of trial spacing has not been listed, since only one relevant empirical study (**12**) is so far available.

All the relationships included have been discussed above except for the last, and this is clearly a direct consequence of the basic assumptions of the theory. In all cases, except perhaps for item *d*, there seems to be unequivocal correspondence of fact and theory. In the case of item *d*, the predicted relationship has been obtained more regularly during extinction than during conditioning. Possibly the difference is attributable to the fact that comparisons of massed and spaced trial groups are better controlled during extinction when such extraneous variables as satiation and habituation of UR's are excluded.

Several additional statements derivable from the theory seem to be almost certainly true, although not yet established by experiment. One of these is the prediction that resistance to extinction will be increased by additional training given after the curve of learning in terms of response probability or latency has reached an asymptote; this relationship seems to have been assumed by Skinner in his formulation of the reflex reserve concept (**21**) but has not, to my knowledge, been subjected to a formal check by experiment. A second example is the prediction that resistance to reconditioning will be directly related to spacing of extinction periods.

The principal set of relevant empirical phenomena not handled at all by the present theory concerns certain properties of response decrement under extremely massed trials. In some experiments with very short intertrial intervals or with conditions of continuous work, as in pursuit rotor performance, asymptotes of intraperiod learning are depressed so that it becomes impossible to achieve the simplifying condition, assumed in the derivations of this paper, that all stimuli available during a period become conditioned to a designated response. In the same experiments reminiscence is commonly observed following rest. These effects apparently are not predictable from the fluctuation theory. The question remains open whether they can be accounted for in terms of variables not so far incorporated into the theory, e.g., response-produced stimulation. For the present these phenomena will provide a source of comfort to those who prefer to terminate analysis at the "molar" level of Hull's (**10**) reactive inhibition construct.

SUMMARY

A mathematical model embodying the concept of random stimulus fluctuation has been found to describe in detail numerous phenomena of spontaneous recovery, regression, or short-term forgetting, and effects of trial spacing at intermediate intervals. This information should be of aid in appraising alternative strategies for theoretical development in this area. It can now be seen that if neo-Pavlovian conceptions of inhibitory potential are to be maintained without major revision, it will be necessary to assume that the amount of random variation in standard experimental situations is negligible. On the other hand, if the amount of random variation is assumed to be appreciable, then a substantial class of distributional phenomena can be accounted for without additional hypotheses.

REFERENCES

1. BARON, M. R. The effect of long intertrial intervals on the limit of eyelid conditioning. *J. exp. Psychol.*, 1952, **44**, 438–441.
2. BUSH, R. R., & MOSTELLER, F. A mathematical model for simple learning. *Psychol. Rev.*, 1951, **58**, 313–323.
3. DEESE, J. *The psychology of learning.* New York: McGraw-Hill, 1952.
4. ELLSON, D. G. Successive extinctions of a bar-pressing response in rats. *J. gen. Psychol.*, 1940, **23**, 283–288.
5. ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
6. ESTES, W. K. Statistical theory of spontaneous recovery and regression. *Psychol. Rev.*, 1955, **62**, 145–154.
7. ESTES, W. K., & BURKE, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.
8. HOMME, L. E. Spontaneous recovery from extinction in relation to number of reinforcements, spacing of acquisition, and duration of initial extinction period. Unpublished doctor's dissertation, Indiana Univer., 1953.
9. HOVLAND, C. I. Human learning and retention. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 613–689.
10. HULL, C. L. *The principles of behavior.* New York: Appleton-Century-Crofts, 1943.
11. LAUER, D. W., & ESTES, W. K. Rate of learning successive discrimination reversals in relation to trial spacing. *Amer. Psychologist,* 1953, **8**, 384. (Abstract)
12. LAUER, D. W., & ESTES, W. K. Observed and predicted terminal distributions of response probability under two conditions of random reinforcement. *Amer. Psychologist,* 1954, **9**, 413. (Abstract)
13. LAUER, D. W., & ESTES, W. K. Successive acquisitions and extinctions of a jumping habit in relation to schedule of reinforcement. *J. comp. physiol. Psychol.*, 1955, **48**, 8–13.
14. MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.
15. MUNN, N. L. *Handbook of psychological research on the rat.* New York: Houghton Mifflin, 1950.
16. OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.
17. REYNOLDS, B. Extinction of trace conditioned responses as a function of the spacing of trials during the acquisition and extinction series. *J. exp. Psychol.*, 1945, **35**, 81–95.
18. ROHRER, J. H. Experimental extinction as a function of the distribution of extinction trials and response strength. *J. exp. Psychol.*, 1947, **37**, 473–493.
19. SCHOEFFLER, M. S. Probability of response to compounds of discriminated stimuli. *J. exp. Psychol.*, 1954, **48**, 323–329.
20. SHEFFIELD, V. F. Resistance to extinction as a function of the distribution of extinction trials. *J. exp. Psychol.*, 1950, **40**, 305–313.
21. SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century-Crofts, 1938.
22. SPENCE, K. A., & NORRIS, EUGENIA B. Eyelid conditioning as a function of the intertrial interval. *J. exp. Psychol.*, 1950, **40**, 716–720.
23. STANLEY, W. C. Extinction as a function of the spacing of extinction trials. *J. exp. Psychol.*, 1952, **43**, 249–260.
24. TOLMAN, E. C., & POSTMAN, L. Learning. In *Annu. Rev. Psychol.*, 1954, **5**, 27–56.
25. UNDERWOOD, B. J. Studies in distributed practice: XII. Retention following varying degrees of original learning. *J. exp. Psychol.*, 1954, **47**, 294–300.

(Received December 22, 1954)

Psychological Review
Vol. 62, No. 5, 1955

RELATIONSHIPS BETWEEN SHAME AND GUILT IN THE SOCIALIZING PROCESS

DAVID P. AUSUBEL

Bureau of Educational Research, University of Illinois

Guilt is one of the most important psychological mechanisms through which an individual becomes socialized in the ways of his culture. It is also an important instrument for cultural survival since it constitutes a most efficient watchdog within each individual, serving to keep his behavior compatible with the moral values of the society in which he lives. Without the aid rendered by guilt feelings, child rearing would be a difficult matter indeed. If children felt no sense of accountability or moral obligation to curb their hedonistic and irresponsible impulses, to conform to accepted social norms, or to acquire self-control, the socializing process would be slow, arduous, and incomplete. Sheer physical force, threat of pain, deprivation, and punishment, or withholding of love and approval would be the only available methods—combined with constant surveillance—to exact conformity to cultural standards of acceptable behavior. And since it is plainly evident that the maintenance of perpetual vigilance is impractical, that fear alone is never an effective deterrent against antisocial behavior, and that the interests of personal expediency are not always in agreement with prescribed ethical norms, a social order unbuttressed by a sense of moral obligation in its members would enjoy precious little stability.

Within recent years, a number of social anthropologists (3, 7, 8, 9) have advanced the notion that guilt is not universally present or prominent as a sanction in mediating and sustaining the culture. Instead they have identified guilt as a unique property of the characterology of individuals who as children experience the kinds of relationships with parents allowing for "superego" formation, that is, typically of persons growing up in cultures adhering to the Judaic-Christian tradition. Thus, among the Samoans, Iatmul, and Balinese, Margaret Mead believes that the culture is primarily transmitted through such external sanctions as expediency in conforming to the rules (8, p. 514), shared, undifferentiated fear (9, p. 369), and anticipation of physical reprisals (9, p. 370), respectively. Ruth Benedict minimizes the importance of guilt and emphasizes the role of shame in regulating the social behavior of the Japanese (3, p. 222). Leighton and Kluckhohn make the same point in reference to the Navaho: "Sensitivity to shame . . . largely takes the place that remorse and self-punishment have in preventing anti-social conduct in white society" (7, p. 106). And even in characterizing the moral development of adolescents in our own society who undergo peer-group rather than parent-regulated socialization, Margaret Mead comments that "shame, the agony of being found wanting and exposed to the disapproval of others, becomes a more prominent sanction behind conduct than guilt, the fear of not measuring up to the high standard which was represented by the parents" (8, p. 520).

In this paper we shall be concerned with a critical examination of the criteria that these anthropologists have used in differentiating between shame and guilt. We shall attempt to show that although the two kinds of sanctions are distinguishable from one another, they are nevertheless neither dichotomous nor mutually exclusive, and that the development of guilt feelings is not dependent upon highly specific aspects of a unique kind of parent-child relationship. Before turning to this task, however, it might be profitable to undertake a logical analysis of the developmental conditions under which the capacity for acquiring guilt behavior arises, as well as some of the basic relationships between shame and guilt. Because of the paucity of experimental or naturalistic evidence in this area of theoretical inquiry, it should be self-evident that the following paradigm is offered as a system of interrelated hypotheses rather than as an exposition of empirically established facts.

THE DEVELOPMENTAL ORIGINS OF GUILT

Guilt may be conceptualized as a special kind of negative self-evaluation which occurs when an individual acknowledges that his behavior is at variance with a given moral value to which he feels obligated to conform. It is a self-reaction to an injured conscience, if by conscience is meant an abstraction referring to a feeling of obligation to abide by all internalized moral values. The injury consists of a self-perceived violation of this obligation. Hence, in accordance with this formulation, one might hypothesize that before guilt feelings can become operative, the following developmental conditions must apply: (*a*) the individual must accept certain standards of right and wrong or good and bad as his own, (*b*) he must accept the obligation of regulating his behavior to conform to whatever standards he has thus adopted, and must feel accountable for lapses therefrom, and (*c*) he must possess sufficient self-critical ability to recognize when a discrepancy between behavior and internalized values occurs.

It goes without saying that none of these conditions can ever be satisfied at birth; under minimally favorable circumstances, however, all human beings should be potentially capable of acquiring the capacity for guilt behavior. Culture may make a difference in the form which this behavior takes and in the specific kinds of stimuli which instigate it; but the capacity itself should be so basically human and so fundamental to the sanctions by which social norms are maintained and transmitted to the young in any culture that differences among individuals within a culture would probably be as great as or greater than differences among cultures.

It is theoretically possible, of course, that in certain extreme cases a culture may be so anarchic and unstructured in terms of the obligations it engenders in its members that the potentiality for guilt experience is never realized, as, for example, among the Dobu (2, p. 131). Ordinarily, however, we might expect that guilt feelings would be found universally; and, hence, the burden of proof regarding their alleged absence in a given culture more properly rests with the investigator making the allegation.

Despite the probable existence of many important culturally conditioned differences in children's acquisition of guilt behavior, there are presumptive grounds for believing that considerable communality prevails in the general pattern of sequential development. Such communality would be a product of various uniformities regarding (*a*) basic conditions of the parent-child relationship, (*b*) minimal cultural needs for socialization of the child, and (*c*) certain gross trends in cognitive and social growth that prevail from one culture to the next.

The cultural basis of conscience development in the individual may be found in the potent need of both parents and society to inculcate a sense of responsibility in the child. Not only the physical survival of its members, but also the perpetuation of its selective way of life is contingent upon the culture's degree of success in this undertaking. Thus, the attenuation of infantile irresponsibility might be considered part of the necessary process of ego devaluation and maturation that presumably characterizes personality development in all cultures. Socialization demands the learning of self-control and self-discipline, the subordination of personal desires to the needs and wishes of others, the acquisition of skills and self-sufficiency, the curbing of hedonistic and aggressive impulses, and the assimilation of culturally sanctioned patterns of behavior. It seems highly unlikely that any of these propensities could become thoroughly stable before conscience is firmly established.

We might postulate that the first step in the child's development of conscience involves his assimilation of parental values and standards.[1] Having no other frame of reference for judgments of good and bad, the prestige suggestion of parents easily holds sway. But acceptance of these values does not obligate him—in the absence of a still undeveloped sense of moral responsibility—to regulate his own behavior accordingly. Lapses on the part of *other* persons are perceived as "bad," but such judgments have no relevance for similar behavior of his *own* when gratification of his *own* impulses is at stake. At this stage of development his behavior can only be directed into acceptable channels by punishment or by anticipation of punishment in the form of pain, deprivation, ridicule, threatened separation from the parent, etc. Conformity to ethical standards, therefore, is "devoid of moral implications because it is only indicative of submission to authority rather than acceptance of it" (1, p. 122).

Behavior can first be regarded as manifesting moral properties when a sense of obligation is acquired. The central hypothesis of the present formulation is that this development typically takes place in children who are accepted and intrinsically valued by parents, and who thereby acquire a derived or vicarious status in consequence of this acceptance. By the fiat of parental acceptance they are provided with intrinsic feelings of security and adequacy despite their manifest dependency and incompetence to fend for themselves. They accordingly become disposed to accept parental values implicitly and unconditionally out of loyalty to the individuals[1] to whom they owe their status and self-esteem. Among other standards assimilated in this uncritical and subservient fashion "is the feeling of moral responsibility or accountability to conform to standards of behavior which have been given ethical implications. . . . Unlike other values which are ends in themselves, moral responsibility has [the] regulatory function of compelling adherence to internalized norms of behavior" (1, p. 135).

It is reasonable to suppose, however, that the inhibition of unacceptable behavior by feelings of personal loyalty and accountability, by the child's recognition of his parent's moral authority, and by his desire to avoid the self-punishing consequences of guilt develops relatively slowly. The old external sanctions of pain, punishment, threat, and ridicule continue to be applied, reinforced now by the more meaningful threat of withdrawal of love and approval. Delaying the consolidation of guilt behavior is the slow growth of both the self-critical faculty (making perception of discrepancies between precept and conduct difficult) and of the ability to generalize principles of right and wrong beyond specific situations.

---

[1] In cultures characterized by an extended family group, or where persons other than parents take major responsibility for child rearing, both value assimilation and the subsequent development of moral obligation have from the very beginning a wider social base.

During preadolescence and adolescence, as the child begins to lose his volitional dependence upon parents and becomes more concerned both with a primary status based upon his own competencies and with equal membership in the social order, the basis of conscience and guilt behavior apparently undergoes significant change. As Piaget has shown, greater experience in interpersonal relationships and in playing differentiated roles in complex social activities makes him more inclined to interpret rules of conduct as functional contrivances designed to facilitate social organization and interaction rather than as sacred and axiomatic givens (12, p. 106). Concomitantly, he develops a notion of moral law based upon principles of equity and embodying a system of reciprocal as opposed to unilateral moral obligations (12, p. 387). The absolutism of his moral standards also tends to be weakened by his greater need for self-enhancement as an individual in his own right —a need that no longer favors unconditional adherence to an uncompromisable set of parental standards, and places a greater premium on the value of expediency. Finally, once the individual acquires this functional concept of moral law, conceives of himself as an independent entity striving for primary status in a social community, and recognizes the reciprocal nature of obligation, he completes the process of transferring "his feeling of moral accountability from parents to the moral authority of society" (1, p. 482).

In heterogeneous cultures, values tend to acquire a wider social base during adolescence. The individual is exposed to a variety of ethical standards and can, within limits, choose between alternative moral systems. And where the peer group plays the major role in adolescent socialization, parents are replaced by peers as the chief interpreters and enforcers of the moral code and, to some extent, as the source of moral authority. But neither set of conditions is essential for the maturational changes in the nature of moral organization that take place during and subsequent to preadolescence (8, p. 519; 10, p. 96). In certain homogeneous cultures also, the source of moral values and authority is referred almost from the beginning to persons outside the immediate family circle (7, p. 51).

ATYPICAL DEVELOPMENT OF GUILT BEHAVIOR

The above description of the original and subsequent development of guilt behavior is hypothesized as the more typical course of moral growth in most children. It presupposes that the latter can achieve derived feelings of status through the medium of a dependent parent-child relationship, and can consequently internalize both values and a sense of moral obligation on the implicit basis of personal loyalty. In all cultures, however, a variable number of parents are psychologically incapable of extending acceptance and intrinsic valuation to their offspring. Thus deprived of the self-esteem derived from the fiat of unconditional parental acceptance, such children are from the very beginning obliged to seek primary status and feelings of adequacy on the basis of their own competencies and performance ability. Accordingly, the basis on which they internalize values and moral obligations might be expected to be correspondingly different.

It would serve no useful purpose here to speculate further on the possible atypical courses that conscience development could take in rejected and extrinsically valued children (1). However, there is less reason to believe that rejecting and extrinsically valuing attitudes characterize *all* parents in a few rare cultures than that they may be found among *some* parents in almost all cultures. Hence, we would hypothesize that the probability of finding some conscienceless individuals in every culture is greater than that of finding conscienceless cultures.

CLASSIFICATION OF SHAME AND GUILT

Generally, shame may be defined as an unpleasant emotional reaction by an individual to an actual or presumed negative judgment of himself by others resulting in self-depreciation vis-a-vis the group. This definition of shame is inclusive of instances in which the persons passing judgment are valued either positively or negatively by the individual being judged (9, p. 367), and of both moral and nonmoral causes for the instigation of such negative judgments. Typical examples of nonmoral shame are embarrassment in committing a breach of propriety or in having one's bodily intimacy exposed to public scrutiny (11, p. 350), and "loss of face" resulting from exposure of ignorance or incompetency. Moral shame, on the other hand, is a reaction to the negative moral judgments of others. It is a prominent sanction in many cultures both before and subsequent to the development of conscience.

Moral shame can, in turn, be divided into two categories—internalized and noninternalized. The latter variety occurs when an individual reacts with self-depreciation to the moral condemnation of others, but does not accept the moral value to which he has failed to conform, e.g., a young child may be shamed by being caught in a lie even if he does not accept the judgment that lying is wrong. Essential for this type of shame is the presence of witnesses to or eventual discovery of the misdeed in question. When the value is internalized, e.g., when the child *really* believes that lying is opprobrious, actual discovery is unnecessary; shame can result merely from presumed or fantasied reproach. Under the influence of such unwitnessed shame, the Ojibway may commit suicide (9, p. 367). According to Margaret Mead, internalized shame only occurs when the parent is the interpreter and enforcer of the sanction (9, p. 366). Observation of closely-knit adolescent peer groups, however, seems to indicate that genuine shame may occur when adolescents perceive or fantasy the moral disapproval of their fellows for offenses against norms established and enforced by the group itself.

The shame associated with guilt may be considered a special case of moral shame. In addition to an internalized moral value, a personal obligation to abide by that value is at stake, e.g., the child not only believes that lying is wrong and feels shame by experiencing or imagining the censure of others, but also feels shame in fantasying the reproach of others for violating a moral obligation. In either case an actual audience is unnecessary, but a presumed or fantasied judgment by others is. However, it is important to emphasize at this point that shame is only one component of guilt, the component involving external judgment and sanction. Guilt also involves other *self-reactions* that are independent of the judgment of others, namely, self-reproach, self-disgust, self-contempt, remorse, lowered self-esteem, anxiety, and various characteristic and subjectively identifiable visceral and vasomotor responses. And conversely, the shame of guilt is only one of many kinds of shame. Thus, the presence of guilt-related shame does not preclude the simultaneous operation of other forms of moral or nonmoral shame unassociated with guilt; nor does the excitation of any of these kinds of shame (associated or unassociated with guilt) preclude the operation of the various self-judgments characterizing guilt feelings.

Shame relies on external sanctions alone. Guilt relies on both internal and external sanctions. The latter sanctions consist of the presumed judgments of others regarding one's lapses with respect to moral obligations and the resulting self-depreciation vis-a-vis the group, as well as the customary social reprisals associated with the misdemeanors arousing guilt. In addition, feelings of guilt have external reference in that they acknowledge accountability for a moral offense against the group.

In the following section we shall critically appraise (*a*) the views of Margaret Mead and of Leighton and Kluckhohn, who hold that the development of guilt behavior requires a parent-child relationship in which an omniscient parent is the referent for and the source of moral authority, in addition to personally administering and interpeting moral sanctions (9, pp. 366, 367); and (*b*) Ruth Benedict's position that guilt is shameless, involves no external sanctions, and is concerned with inner convictions of sin (3, pp. *222–223*).

ETHNOCENTRIC DISTINCTIONS BETWEEN SHAME AND GUILT

Superego Models of Guilt

The first ethnocentric conception of guilt referred to above equates the capacity for experiencing guilt feelings with neo-Freudian notions of superego formation. According to this conception, guilt feelings cannot arise unless the conditions essential for superego formation are not only present in childhood but are also maintained in subsequent years. Hence, guilt-like behavior in our own or other cultures that does not conform to these conditions is categorized as shame. The two basic conditions that are laid down for the development of guilt are: (*a*) the child must accept the parent or parent surrogate as omniscient or as qualitatively superior to himself in a moral sense, and (*b*) he must accept the parent as the source of moral authority, i.e., as the "referent" in whose name moral behavior is enjoined.

In reference to the former condition Margaret Mead states:

> The adults in the society must think of the child as qualitatively different from themselves, in that the child has not yet attained their moral stature, but is subject to innate impulses which, if permitted unchecked expression, would eventuate in adult character different from and morally inferior to that of the parent (8, p. 514).

Leighton and Kluckhohn similarly characterize guilt feelings as representing an individual's acknowledgment that he is "unworthy . . . for not living up to the high standards represented by [his] parents" (7, p. 106). In explaining why "Navaho sensitivity to shame takes the place of guilt in our society," they go on to say:

> "Conscience" is related to the belief in an omnipotent God who knows all. . . . In white society the child doesn't see the parent in many life situations. [But] in the circumstances of Navaho life a pose of omnipotence or omniscience on the part of parents would be speedily and almost daily exposed (7, p. 106).
>
> Only rarely does one hear the utterances on the part of Navaho parents which are so usual among white parents: "Do it because I say it is right," "do it because I say so," "do it because I am your father and children must obey their parents" (7, p. 53).

In the first place, one may legitimately question the assumption that children do not acquire moral obligations unless a highly authoritarian, paternalistic, and hierarchical parent-child relationship exists, unless the parent sets himself up and is perceived as a qualitatively superior paragon of virtue. As long as the conditions for acquiring derived status are fulfilled, it seems reasonable to suppose that moral obligations are incurred provided that parents or parent surrogates make such expectations unambiguously clear. It undoubtedly increases the prestige suggestion with which the standards and expectations of parents are endowed if the latter individuals are perceived as omniscient; but there is neither convincing logical nor empirical justification for designating this criterion as a *sine qua non* for the acquisition of guilt behavior.

A second difficulty confronting the proposition that qualitative hierarchical distinctions must prevail between preceptor and learner in order for guilt feelings to develop is the suggestive evidence of Piaget, referred to above, showing that as children approach adolescence in our culture they become increasingly concerned with moral obligations based on principles of reciprocity (12, pp. 106, 387). Cooperative and functional relationships with peers engender notions of obligation reflecting mutual respect between equals which gradually displace unilateral feelings of obligation based on implicit acceptance of authority (12, p. 106).

The second condition imposed by adherents of the superego model of guilt behavior, i.e., that parents or parent surrogates must serve as the source of moral authority, runs into similar difficulties posed by developmental shifts in the organization of conscience. It presupposes, in the words of Margaret Mead, that

> . . . teachers, the clergy, judicial officials, etc., partake of the character of the judging parent who metes out reward to the individual child (or, later, adult) who makes a satisfactory approximation to the desired behavior, and punishment to him who does not (8, p. 514).

Such a requirement arbitrarily eliminates the possibility of guilt behavior developing in cultures like the Navaho where the authority for moral sanctions is from the earliest days of socialization referred to the group as a whole (7, p. 51). The same criterion rules out the occurrence of guilt behavior among (*a*) adolescents in our own culture who accept the peer group as their major source of moral authority (see earlier quotation from Margaret Mead), (*b*) rejected and extrinsically valued children who never accept the moral authority of parents, but who may accept rational principles of equity and the moral authority of the group, and (*c*) adults in most cultures, who, in accordance with the developmental changes described above, transfer their allegiances from the moral authority of parents to the moral authority of society. The latter process does not necessarily require the intermediate adolescent step of repudiating adult authority and substituting peer standards. Sufficient for this outcome is a transformation of the basis upon which the authority of elders is accepted, i.e., a change from implicit acceptance of axiomatic truth by a dependent and ideationally subservient individual to a more rational acceptance of functionally necessary norms growing out of interpersonal relations by an independent member of adult society (10, pp. 95–96).

To deny that guilt feelings can arise in the above four situations requires summary dismissal of the evidence of everyday experience with the behavioral content of others. It is admittedly difficult to evaluate the subjective emotional experience of other persons, particularly if they happen to be members of different cultures. Logical inference must always supplement the perceptions of even the most experienced observer. Nevertheless a decision regarding the meaning of behavior can never be reached on the basis of a priori criteria alone, especially when these lack irrefutable logical or face validity. An emotion is adjudged to be guilt or shame on the basis of its behavioral characteristics, its situational excitants, and its subjective properties, and not because it fails to meet certain arbitrary ethnocentric criteria that apply to a single culture or phase of development. In the next section we shall examine some ethnological materials presented by Leighton and Kluckhohn and by Ruth Benedict as illustrative of shame and show how such behavior can be more defensibly interpreted as guilt.

A conception of conscience or guilt behavior that makes no provision for developmental changes in the underlying personality structure upon which such behavior depends cannot be successfully integrated into any self-consistent theory of personality development. A theoretical framework must be found that is broad enough to encompass the guilt behavior of a child of eight and likewise of an adult of eighteen or eighty. We cannot conclude that eighteen-year-old guilt must be shame because it does not conform to an eight-year-old model of guilt.

Personality development is admittedly characterized by continuity as well as by change. A tendency to accept values and obligations uncritically, on the basis of personal allegiances, still continues in adult life. Similarly, early attitudes of unconditional loyalty to moral obligations tend to persist long beyond the span of childhood. Nevertheless although these continuing trends add much to the stability of conscience, they are largely supplanted by different moral orientations and enjoy only substrate existence in a generally changed gestalt. A precipitate of the experiences of the eight-year-old is represented in the behavior of the eighteen-year-old, but this precipitate has a phenomenologically different meaning and significance than its original form a decade earlier.

Pointing up the logical as well as the empirical untenability of a criterion of guilt that requires a qualitatively superior parent or parent surrogate as the source of moral authority is the discrepancy on this score between the two chief examples of allegedly "shaming cultures," i.e., certain North American Indian tribes and the Japanese. Leighton and Kluckhohn characterize the Navaho as a "shaming culture" in part because sanctions are applied laterally rather than from above, and are referred to group rather than to parental authority (7, pp. 105–106). On the other hand, Ruth Benedict, using other criteria, regards the Japanese as a shaming culture despite the fact that the Japanese family is the ethnological example par excellence of unconditionally accepted, paternalistic (if not arbitrarily imposed) authority based upon hierarchical position (3, pp. 55–57, 264). Obviously, then, this criterion for distinguishing between shame and guilt has little claim to universality.

Guilt as a Shameless, Wholly Internalized Conviction of Sin

The second major ethnocentric (and psychoanalytically oriented) conception of guilt assumes that guilt and shame are mutually incompatible and that genuine guilt must be devoid of shame and all external sanctions. Thus, in cultures where shame plays a more prominent role in behavioral control than in our own culture, ethnologists sharing this conception of guilt have insisted that guilt feelings must *ipso facto* be either nonexistent or of negligible significance. Among the Japanese, according to Gorer, "mockery is the most important and most effective sanction for obtaining conformity in social life" (4, p. 20). Ruth Benedict espouses the same opinion in stating that "shame has the same place of authority in Japanese ethics that 'a clear conscience,' 'being right with God,' and avoidance of sin have in Western ethics" (3, p. 224). Leighton and Kluckhohn claim that Navahos do not internalize the "standards of their parents and elders," and that shame ("I would feel very uncomfortable if anyone saw me deviating from accepted norms") is the major deterrent to antisocial behavior (7, p. 106). Margaret Mead goes even further by ruling out the occurrence of both guilt and shame in the sanction systems of the Samoans (8, pp. 514–515), Balinese (9, p. 369), and Iatmul (9, pp. 369–370), and insisting that only such external sanctions as fear, expediency, and threat of physical reprisals are operative.

This dichotomization of shame and guilt, of internal and external sanctions, in addition to lacking logical self-consistency, simply does not conform to available naturalistic evidence. Both moral and nonmoral shame may be extremely important sanctions in a given culture, but they preclude the occurrence of neither the public (shame) nor the private (self-judgment) aspects of guilt. A culture so unstable and so anarchic that its members fail to respond to internal sanctions (i.e., feelings of moral obligation) is indeed an ethnological rarity. The inhabitants of Dobu (2) may fit this description, but the case made out for the Samoans, Iatmul, and Balinese (8, 9) is far from convincing; and descriptions of Japanese (3, 6) and Navaho (7) behavior tend to disprove the very contentions they are meant to illustrate. Conversely, a culture in which external sanctions are *not* applied at any stage of moral development is also rare enough to qualify as an ethnological oddity.

The evidence seems clear enough that although Japanese and Navaho individuals are more responsive to shame than we are, much of this shame is really the shame of guilt, i.e., the shame accompanying awareness of violated moral obligations. The predominant pattern of parent-child relationships in both cultures is one of acceptance and intrinsic valuation of the child, leading to the first or implicit stage of moral responsibility based upon personal loyalty (3, Ch. 12; 7, Ch. 1). Navaho children acquire a sense of responsibility between five and eight years of age, learning assiduousness in executing their chores. They learn "that they cannot always indulge themselves and that they have duties toward others" (7, p. 58). Navaho parents are strikingly devoted and attentive to the needs of their children (7, pp. 13–33), and children, in turn, even as adults, are very devoted to their parents and siblings (7, pp. 94–100). The Navaho adult is capable of very responsible conduct in relation to his social group despite the fact that responsibility is "divided and shared" (7, p. 107). If such behavior is not indicative of internalized moral obligation, what other credible explanation is there? Surely shame alone could not account for all of it.

Among the Japanese an even stronger case can be made for the importance of conscience and guilt behavior. Few other cultures lay greater stress upon the sanctity of moral obligations. "The strong according to Japanese verdict are those who disregard personal happiness and fulfill their obligations. . . . A man is weak if he pays attention to his personal desires when they conflict with the code of obligations" (3, pp. 207–208). Obligations to others must be fulfilled even if one dislikes the latter personally (3, p. 124). The Japanese sense of accountability is especially stringent. "Each man must accept responsibility for the outcome of his acts. He must acknowledge and accept all natural consequences of his weakness, his lack of persistence, his ineffectualness" (3, p. 296).

What is even more impressive is the fact that in contrast to prevailing attitudes in our own culture, the self-discipline, the self-restraint, and any of the personal disadvantages incurred in honoring obligations are not regarded as self-sacrifice. The fulfillment of obligations is taken for granted on the grounds of reciprocity, and calls for no applause, reward, self-pity, or self-righteousness (3, pp. 230–233). The notions of mutual dependence and indebtedness to one's society and one's past lie at the very heart of the Japanese concept of obligation. "One repays one's debts to one's forbears by passing on to one's children the care one oneself received" (3, p. 122).

Despite this overwhelming evidence of conscience and guilt behavior, Ruth Benedict is misled by the dichotomized conception of shame and guilt, and insists upon referring to such behavior as shame on the grounds that guilt is supposedly shameless. But, as already pointed out above, what she calls "shame" is actually the shame component of guilt. La Barre paints essentially the same picture of Japanese character structure as she does and reaches a diametrically opposite conclusion. He asserts that the Japanese superego is exceedingly strong (6, p. 329). Nevertheless, even while minimizing the role of guilt among the Japanese, Ruth Benedict is grudgingly obliged to admit the existence of moral self-judgments unrelated to external sanctions.

> They [the Japanese] are terribly concerned about what other people will think of their behavior, and they are also overcome by guilt when other people know nothing of their misstep. . . . Japanese sometimes react as strongly as any Puritan to a private accumulation of guilt (3, pp. 2–3, 222).

We cannot escape the conclusion, therefore, that both guilt and shame, and internal and external sanctions, can and do exist side by side and mutually reinforce each other. The assertion that "true shame cultures rely on external sanctions for good behavior, not, as true guilt cultures do, on an internalized conviction of sin" (3, p. 323) is unsupported by available evidence. The presence of the stock, the pillory, and the ducking stool in the public market place offers eloquent refutation to the statement that "the early Puritans who settled in the United States tried to base their whole morality on guilt" (3, p. 223). Reinforcing most of the moral sanctions that we customarily assign to the domain of conscience is a parallel set of statutes and group pressures enforced by appropriate public reprisals. Even in cultures where moral obligations are highly internalized, we usually find a policeman on the corner giving a friendly nudge to sluggish consciences or a timely warning to impish consciences pondering a brief vacation from duty.

The regulation of conduct by moral self-judgment is also the *final* step in a long developmental process. In controlling the behavior of children in our own culture, parents may differ from Samoan, Balinese, and Iatmul parents in their propensity for premature sermonizing about right and wrong, filial duty, virtue, etc. But during the first few years of life such preachments are but empty verbalisms to our children, who actually respond as do children in all cultures to the accompanying external sanctions of reward and punishment. Guilt feelings also have external reference by virtue of their shame component which involves a self-deprecatory reaction to the presumed or actual moral censure of others for a transgression of moral obligation. Finally, the external reference of guilt is apparent in the fact that it is reduced by punishment and confession. It always implies an offense against the group which, therefore, can only be pardoned by group action.[2]

Another source for ethnocentric misinterpretation of Japanese guilt behavior lies in their lack of preoccupation with sin. The idea of sin, however, is in no way indigenous to a conscience founded on moral obligation; it merely tends to be associated with a specific variety of guilt behavior that is influenced by certain religious traditions and notions about the original nature of man. Nevertheless, by equating guilt with "an internalized conviction of sin" (3, p. 223), and conscience with "the avoidance of sin" (3, p. 224), Ruth Benedict finds further "justification" for regarding the Japanese as relatively guiltless.

Japanese neglect of the problem of sin, however, stems from two other sources that can hardly be interpreted as detracting from the genuineness of their expressions of guilt. In the first place, since they do not regard indulgence in physical pleasure as inherently evil—providing that it is properly subordinated to moral obligation—they do not share our notion of "original sin" or our conception of human nature as constantly in need of redemption from inherently sinful desires (3, pp. 191–192). Secondly, since the fulfillment of obligation is taken for granted and is not regarded as self-sacrifice or as a victory of virtue over innately evil propensities, Japanese good conscience is not burdened with the self-righteousness that comes from conquering sin; and, similarly, guilt is not "internalized conviction of sin," but simply awareness of the breakdown of moral obligation.

Because the Japanese lay great stress on certain "external" (shame) aspects of self-respect, such as acting with circumspection and paying scrupulous attention to the details of propriety, Western observers are prone to deduce that this precludes their association of the term *self*-respect with a criterion such as "consciously conforming to a worthy standard of conduct" (3, p. 219). However, in the light of the foregoing description of Japanese concepts of moral obligation, we are only justified in concluding that *in addition* to sharing the meaning which is commonly placed on "self-respect" in our culture, the Japanese value other character traits (i.e., self-restraint, circumspection, respect for social niceties and for the feelings of others) sufficiently to dignify them by the same term that we usually reserve for moral uprightness.

Lastly, it is important to explain why confession and atonement are such relatively inconspicuous aspects of Japanese guilt behavior.[3] The absence of these characteristics that so frequently accompany the expression of guilt in our culture does not prove, as Ruth Benedict suggests, that genuine guilt is lacking, since otherwise it would be relieved by confession (3, p. 223). The Japanese does not, as alleged, avoid confession because his guilt is really only shame and, hence, would be aggravated rather than eased by being made public. The more credible explanation is that although confession would be guilt reducing it would also be too traumatic in view of the tremendous Japanese sensitivity to shame. Under such circumstances, overwhelming feelings of guilt can be relieved less painfully by suicide.

---

[2] In private confessionals, the priest serves as the moral representative of the group.

[3] The practice of atonement is not entirely unknown in Japan. "Boys in later elementary school are sometimes confined to the house for kinshin, 'repentance,' and must occupy themselves with . . . the writing of diaries." Similarly, when ostracized by his schoolmates for offenses, the culprit "has to apologize and make promises before he is readmitted" (3, pp. 273–274).

SUMMARY AND CONCLUSIONS

The capacity for experiencing guilt behavior (i.e., a constellation of negative self-reactions to the perceived violation of moral obligation) is conceived as so basically human and so fundamental to the sanctions by which social norms are maintained and transmitted to the young that under minimally favorable social conditions it should develop in all cultures. The only psychological conditions hypothesized as essential for the development of guilt behavior are (*a*) the acceptance of moral values, (*b*) the internalization of a sense of moral obligation to abide by these values, and (*c*) sufficient self-critical ability to perceive discrepancies between internalized values and actual behavior. The development of conscience typically embodies various sequential changes paralleling shifts in the individual's biosocial status. Ordinarily, greater differences in moral behavior may be anticipated among individuals within a culture than among different cultures.

Shame is a self-deprecatory reaction to the actual or presumed judgments of *others*. It may have moral or nonmoral reference. Guilt feelings always involve a special type of moral shame in addition to other negative *self*-judgments. In addition to the shame of guilt, conscience (guilt behavior) is customarily buttressed by such external sanctions as statutory law, public opinion, and ridicule. Thus, the sanctions of guilt are both external and internal in nature.

Various criteria have been advanced by different ethnologists for distinguishing between shame and guilt. It has been alleged (*a*) that guilt behavior can only arise when a hierarchically superior parent or parent surrogate serves as the source of moral authority, (*b*) that genuine guilt feelings can only exist when shame and other external sanctions are not operative, and (*c*) that, subjectively, guilt must be characterized by conviction of sin and need for atonement. Examination of Navaho and Japanese ethnological materials presented in support of these criteria shows that the latter lack both logical and empirical validity. Highly suggestive evidence of the operation of strong moral obligations and of guilt behavior was found in these supposedly prime examples of "shaming cultures."

The problem of shame and guilt is illustrative of a major methodological hazard in the cross-cultural investigation of personality traits, namely, of the ethnocentric tendency to define a given trait in terms of its *specific* attributes in one's own culture. This leads to a perversion of cultural relativism. Instead of demonstrating how a basic human capacity (e.g., guilt behavior) occurs under different conditions and assumes different forms in different cultures, this approach inevitably "discovers" that this very capacity is absent in other cultures because it does not conform to a specific set of ethnocentric crtieria. The generalization is then made that the capacity is nonuniversal, culture-bound, and, hence, not basic. Cross-cultural investigations of personality must, therefore, start with definitions of traits that are both psychologically meaningful and general enough to encompass specific cultural variants.

It is also important to avoid the error of circularity—setting up *a priori conditions* for the emergence of a trait, finding these conditions absent in a given culture, and concluding "empirically" therefrom that the *trait itself* is absent. This error can be avoided in part by paying attention to the subjective and expressive content of behavior and not merely to the postulated determinants of the trait in question.

REFERENCES

1. AUSUBEL, D. P. *Ego development and the personality disorders.* New York: Grune & Stratton, 1952.
2. BENEDICT, RUTH. *Patterns of culture.* Boston: Houghton Mifflin, 1934.
3. BENEDICT, RUTH. *The chrysanthemum and the sword.* Boston: Houghton Mifflin, 1946.
4. GORER, G. *Japanese character structure.* New York: Institute for Intercultural Studies, 1942.
5. JENKINS, R. L. Guilt feelings—their function and dysfunction. In M. L. Reymert (Ed.), *Feelings and emotions.* New York: McGraw-Hill, 1950.
6. LA BARRE, W. Some observations on character structure in the Orient: the Japanese. *Psychiatry,* 1948, **8**, 319–342.
7. LEIGHTON, DOROTHEA, & KLUCKHOHN, C. *Children of the people: the Navaho individual and his development.* Cambridge, Mass.: Harvard Univer. Press, 1947.
8. MEAD, MARGARET. Social change and cultural surrogates. In C. Kluckhohn & H. A. Murray (Eds.), *Personality in nature, society and culture.* New York: Knopf, 1949.
9. MEAD, MARGARET. Some anthropological considerations concerning guilt. In M. L. Reymert (Ed.), *Feelings and emotions.* New York: McGraw-Hill, 1950.
10. NICHOLS, C. A. *Moral education among North American Indians.* New York: Teachers Coll., Columbia Univer., 1930.
11. NUTTIN, J. Intimacy and shame in the dynamic structure of personality. In M. L. Reymert (Ed.), *Feelings and emotions.* New York: McGraw-Hill, 1950.
12. PIAGET, J. *The moral judgment of the child.* Glencoe, Ill.: Free Press, 1932.

(Received January 17, 1955)

Psychological Review
Vol. 62, No. 5, 1955

NOTES TOWARD A MOTOR THEORY OF VISUAL EGOCENTRIC LOCALIZATION[1]

JAN H. BRUELL AND GEORGE W. ALBEE

Western Reserve University and Highland View Hospital, Cleveland, Ohio

The authors are engaged currently in a series of studies of space perception in hemiplegic patients. The known relationship between space perception and tonic processes provided the rationale for our studies with hemiplegics whose main problem, by definition, is a disturbance of motor processes. In the process of collecting and analyzing our data we have found that a variety of phenomena of space and movement perception can be ordered, using a limited number of hypotheses.

Space perception refers to impressions of the positions of objects relative to each other and relative to the observer. The convention in present-day treatment of the problem is to refer to judgments of the relative position of objects to each other as *relative localization* and to judgments concerning the position of objects relative to the observer as *egocentric localization.* The present paper will be restricted to a consideration of egocentric localization, and more specifically to the visual perception of direction and egocentric apparent movement.

VISUAL EGOCENTRIC LOCALIZATION AND THE INNERVATION OF THE EXTERNAL EYE MUSCLES

Retinal Local Signs and Voluntary Eye Movements

The perception of the egocentric position of visual objects is often attributed to a hypothetical function of retinal receptors. Each retinal point is assumed to have a definite space value or local sign. Conventionally (14) receptors to the right of the median meridian of the retina are given negative space values, and those to the left of the meridian positive space values; receptors in the fovea and median meridian are assumed to have a zero space value, as shown in Fig. 1*a*. In this and succeeding figures it is assumed that the stimulus objects are placed at least six feet from the observer. At this distance convergence is negligible and the binocular situation can be represented by a "cyclopean eye." Objects A and B, which stimulate receptors with negative space values, are perceived as being to the observer's left; objects D and E, stimulating receptors with positive space values, are perceived as being to the right; and object C, projected on the fovea (F), is perceived as being straight ahead.

Obviously this applies only when the observer looks at a point straight ahead, as in Fig. 1*a*. In Fig. 1*b*, if the eye is turned *voluntarily* from C to E, then object C, placed in the objective median plane of the observer, stimulates a peripheral part of the retina. Yet the observer continues to perceive C as being straight ahead of him. That is, a peripheral point on the retina has assumed a space value of zero. On the other hand, object E, placed to the right of the objective median plane, now stimulates the fovea but is perceived as being to the right. Clearly, with voluntary eye-turning, retinal points change their space values. With voluntary eye-turning to the right (Fig. 1*b*), the fovea

---

[1] This paper will ultimately be part of a larger study of space perception made possible by a grant-in-aid of research from the Cleveland Foundation.

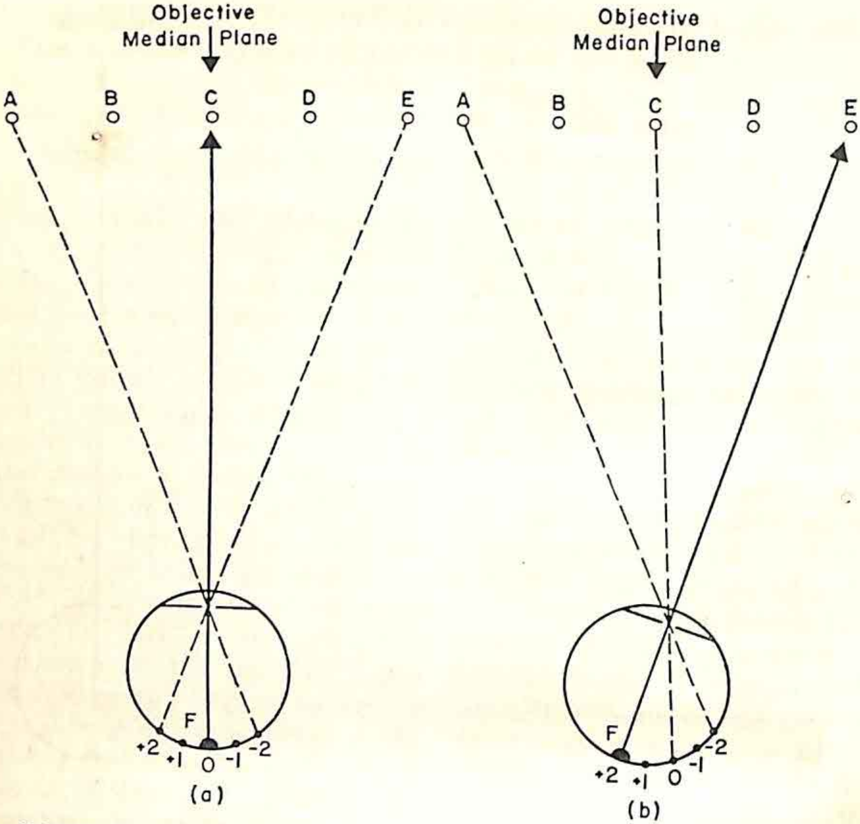


FIG. 1. Schema illustrating concept of retinal space values: (*a*) in primary position of the eye, (*b*) with the eye turned voluntarily to the right.

assumes a positive space value and all other retinal points change their space values correspondingly.

This common observation leads to our first hypothesis: whenever the dextro- or levorotators of the eyes receive *voluntary* innervation, the fovea assumes a positive or a negative space value, respectively, and concomitant changes occur in the space values of all retinal points (Hypothesis 1).

Retinal Local Signs and Involuntary Movements

Figure 2 represents a situation in which the eyes of an observer remain stationary while an object C is moved to the left. Object C stimulates retinal points with space values of − 1, − 2, −3, etc., in succession. Under these conditions C is perceived as moving to one's left, as it actually does. The same retinal stimulation occurs when an object, C, remains stationary while the eyes are turned to the right. However, whether such retinal stimulation is experienced as movement of C to the left depends upon the voluntary or reflex nature of the eye movement.

As was shown above, in the case of *voluntary* eye-turning, retinal receptors change their space values; and C, which stimulates more and more peripheral points on the retina, continues to be perceived as straight ahead. On the other hand, if the eyes turn *involuntarily* to the right, the stationary object C will appear to move to the left. If, for instance, a subject is rotated counterclockwise around his vertical axis, his eyes will move involuntarily to the right (slow phase of vestibular nystagmus). Object C, placed straight ahead of the subject and rotated with him, will stimulate increasingly peripheral points on the right half of the retina of the involuntarily turning eye. These points have negative space values when the eye is in its primary position. From the fact that under the described conditions the object is seen moving to the left, we conclude that with involuntary eye-turning, retinal points do not change their space values.

Thus we arrive at a hypothesis supplementary to the hypothesis derived from our discussion of the effects of voluntary eye-turning on retinal local signs. There we hypothesized that whenever eye muscles receive voluntary innervation, all retinal receptors, including the fovea, change their space values. The fovea assumes a positive space value if one attempts to turn the eyes to the right, a negative space value if the voluntary effort is toward turning the eyes to the left. Now we hypothesize: when eye position is determined only by *involuntary* innervation of the eye turners, retinal receptors do not change their space values (Hypothesis 2). For instance, the fovea retains its space value of zero, and any object projected upon the fovea appears to be straight ahead.

These hypotheses, if valid, should permit us to predict the apparent egocentric position of objects from a knowledge of the pattern of oculomotor innervation obtaining in a given situation. We shall now examine certain factors determining the innervation of the eye turners.

Some Factors Determining Eye Position

The position of the eye in its orbit is determined by the innervation, and consequently by the state of contraction of the six external eye muscles. This innervation may be either voluntary or reflex. The reflex innervation of the eye turners can be brought about by retinal stimulation or it can have its origin in the motor system.

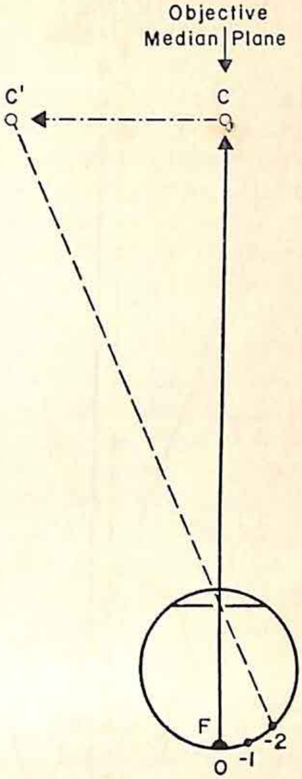


FIG. 2. Schema illustrating retinal stimulation by object moving to the left.

Stimulation of receptors in the periphery of the retina elicits the fixation reflex. The eye turns reflexly in the direction of the stimulus. The reflex eye movement ceases when the stimulus projects on the fovea. Figure 3*b* represents an instance of unilateral retinal stimulation. Stimuli are placed only to the right of fixation, causing reflex eye-turning to the right. As the eye turns, retinal stimulation becomes bilateral as receptors on both sides of the fovea are stimulated. Such bilateral stimulation results in reflex innervation of both the levo- and the dextrorotator muscles.

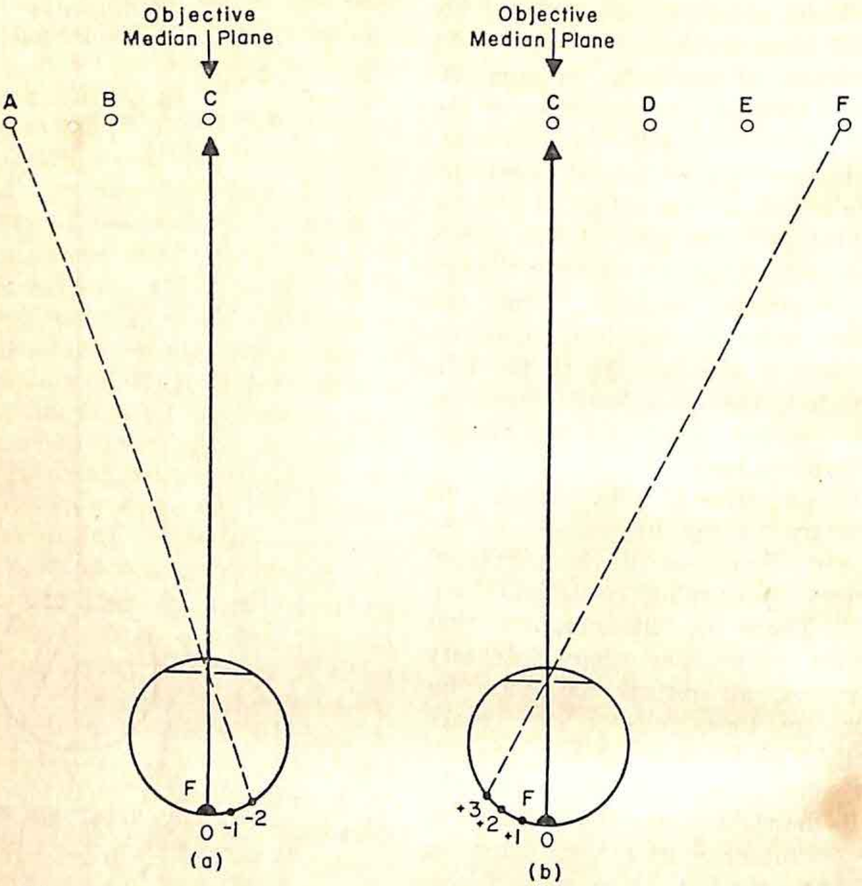


FIG. 3. Schema of visual field extending (*a*) to the left and (*b*) to the right of fixation.

Initially retinal stimulation is asymmetrical, causing stronger innervation of the dextrorotator. When the eye fixates the center of the stimulus field, the fixation reflex will cause innervation of the dextro- and the levorotator, and the eye will cease turning to the right (3).

This analysis assumes that eye position is determined by retinal stimulation only. Nonvisual muscular factors, however, co-determine eye position. In the absence of visual stimulation the eyes tend to assume their "primary position," that is, they tend to look straight ahead. Although it is certain that several reflex mechanisms contribute to this effect, a myotatic reaction is probably involved. When, for instance, the eyes are turned to the right, the levorotators are stretched. This stretching stimulus is thought to elicit a myotatic reflex in which the stretched levorotators contract, thus returning the eyes to their primary position. The recent discovery of kinesthetic receptors in the eye muscles (20, 23) makes the existence of such a myotatic reflex quite plausible. However, as will be seen, whether the eyes are maintained in the median position by this particular reflex or by another reflex mechanism is not crucial to our formulation. In the situation represented in Fig. 3*b*, the fixation reflex acting alone would cause the eyes to turn to the right until the center of the stimulus field is fixated. The myotatic reflex acting upon the stretched levoturners, however, will tend to arrest the turning of the eyes to the right before the center of the stimulus field is fixated.

It is important to note that if the observer is instructed to fixate continuously object C (Fig. 3*b*), he can do so only by counteracting the reflex tendency of the eyes to turn to the right. He must innervate voluntarily the levorotators.

In short, if the position of the eyes is known for a given stimulus situation, it is possible to infer whether eye position is being maintained by reflex innervation of the eye turners alone, or whether it is co-determined by voluntary innervation. Analyses such as these, in conjunction with our hypotheses concerning the effects of voluntary and reflex innervation of eye turners on retinal local signs, will form the basis for our discussion of a variety of phenomena of egocentric localization.

THE PERCEPTION OF THE MAIN COORDINATES OF EGOCENTRIC SPACE: APPARENT STRAIGHT-AHEAD AND APPARENT EYE LEVEL

The Dietzel-Roelofs Phenomenon

Dietzel (5) and, independently, Roelofs (1[illegible]) performed experiments which fit exa[illegible] the paradigm represented in Fig. 3. A luminous figure, the only visible object in a darkroom, was placed in such a position that one of its edges was in the objective median plane of the observer. The observer had the task of judging the egocentric position of C (Fig 3*b*). Our hypotheses permit us to predict the outcome of such an experiment.

It was shown in the preceding section that continuous fixation of C would be possible only if the levorotators received voluntary innervation. Thus, from our hypothesis concerning the effects of voluntary innervation on the local sign of the fovea (Hypothesis 1) it follows that C would be perceived as placed to the observer's left. The experience of the egocentric position of C would be the same, but the underlying physiological processes would differ if the observer were not instructed to fixate C. In this case his eyes would turn reflexly to a point somewhere short of the center of the stimulus field, say to point D. At D, eye position would be maintained, due to the balanced reflex innervation of the levo- and dextrorotators and, according to our hypothesis concerning the effects of involuntary innervation on retinal local signs (Hypothesis 2), the fovea would retain its space value of zero. The foveally projected point D would be seen by the observer as straight ahead of him, while point C, *relative to D,* would appear to be to his left (relative localization; see introduction).

Similarly, if C were at the right extreme of a stimulus configuration, and if it were placed objectively straight ahead of an observer (Fig. 3*a*), it would appear to him to be to his right.

Dietzel (5) and Roelofs (19) actually obtained these results, and they have been confirmed in numerous recent experiments (3, 18, 21). The general findings of these experiments can be accounted for by our hypotheses. The apparent straight-ahead always tends to shift in the direction of the center of the stimulus field. If this center deviates from what is objectively straight ahead of the observer (Fig. 3), objective and subjective straight-ahead do not coincide.

Ordinarily we are fairly accurate in judging what is straight ahead. That is, objective and subjective straight-ahead *do* coincide. This can be accounted for in the same terms as the described discrepancies between objective and subjective space. Usually the whole visual field is filled with stimuli, and thus if a point in the objective median plane is fixated, retinal stimulation tends to be bilateral and symmetrical. This results in equal reflex innervation of the levo- and dextrorotators, and eye position is maintained by reflex innervation alone. Under these conditions, according to our second hypothesis, the fovea will have a space value of zero, and the centrally located fixated object will appear to be straight ahead.

A situation outside of the laboratory in which the center of the stimulus field deviates from what is objectively straight ahead occurs during night driving. In Fig. *3a,* point C might be one of the headlights of an oncoming car. This headlight will be perceptually displaced in the direction of the observer. It is probably a benign illusion as it makes him steer his own car farther away from the passing car than would be required by the objective situation.

Egocentric Localization in Hemianoptics

In the situation diagrammed in Fig. *3b*, a place near the center of the stimulus field appears to be straight ahead. This deviation of what is apparently straight ahead from what is objectively straight ahead can be produced in anyone in a darkroom. In this situation we know that if the subject fixates C (Fig. *3b*), only his left occipital lobe will receive stimulation. The situation is comparable to what happens in a hemianoptic where there is a loss of the left half of the visual field. In the darkroom experiment only the left occipital lobe *receives* stimulation, owing to the experimental arrangement. In the hemianoptic with a loss of the left half of the visual field, only the left occipital lobe *responds* to stimulation, because of the injury in the other half of the visual system. Thus one would expect that in such patients apparent straight-ahead would deviate from objective straight-ahead in their daily perception. This expectation is borne out by clinical observation (2) and experiments (7).

The observation that egocentric localization is disturbed in hemianoptics might lead to the conclusion that the perception of the main space coordinates is linked to purely sensory (occipital) visual processes. Therefore, it should be stressed that our interpretation, which relates phenomena of egocentric localization to processes of oculomotor innervation, is in itself in no way incompatible with the observations made in hemianoptics. On the contrary, as we believe that visual egocentric localization is partially determined by optically induced reflex movements of the eye, and as we know that in man the motor branch of such reflexes has its origin in the occipital lobes (15), we would, on these grounds alone, expect egocentric localization to be disturbed in hemianoptics.

The Robert MacDougall Phenomenon

Our discussion so far has centered around phenomena of egocentric localization in the left-right dimension. What applies to the perception of mediality applies also to the subjective horizon. In fact, it was in the up-down dimension that displacements of objects analogous to those observed by Dietzel and by Roelofs were studied first. Robert MacDougall (16) found at Harvard in 1903 that the fixated upper edge of a stimulus figure, if placed at the objective eye level of an observer, appeared to be above eye level. MacDougall's finding has been confirmed in a number of experiments (10, 11), and can be explained in the same terms as the results of Dietzel and of Roelofs. Fixation of the upper edge of a stimulus figure can be maintained only by the voluntary counteraction of the reflex tendency of the eyes to turn downward toward the center of the stimulus (fixation reflex). The elevator muscles of the eyes have to receive voluntary innervation. This changes the local sign of the fovea from eye-levelness to upness; and the fixated upper edge of the configuration appears to be above eye level. MacDougall's darkroom experiment was inspired by the fact that, in spite of the curvature of the earth, an observer standing on the deck of a ship will perceive the horizon as being above eye level.

APPARENT EGOCENTRIC MOVEMENT

The preceding discussion dealt with the perception of the egocentric position of objectively and phenomenally stationary objects. In what follows we will discuss phenomena of apparent egocentric movement, i.e., we will be concerned with apparent changes in the egocentric position of stationary objects.

If [illegible] observer fixates a stationary light C (Fig. *3b*) and if another light, D, is turned on, he will have a tendency to turn his eyes toward the peripheral light D. In order to counteract this tendency he must voluntarily innervate his levorotators. When light E is turned on, a further increase in the voluntary innervation of the levorotators becomes necessary. The same effects occur when lights F and G are turned on in succession. The increasingly strong reflex innervation of the dextrorotators due to stimulation of more and more peripheral parts of the retina (3) must be counteracted by increasing correspondingly the voluntary innervation of the levorotators.

From our first hypothesis it can be predicted that such a gradual increase in the voluntary innervation of the levorotators would change the space value of the fovea from 0 to − 1, to − 2, to − 3, etc. (cf. above, p. 392). The fixated point C and all other points of the stimulus configuration would appear to move to the left. This prediction is borne out by experiments which we have conducted and which will be published elsewhere.

It is important to note that in this latter situation both object and eye remain stationary. According to the foregoing analysis, apparent movement of the stimulus field will occur if voluntary innervation of one set of eye turners (e.g., the levorotators) increases gradually but does *not* result in actual eye movements. The apparent movement of the stimulus field under such conditions occurs in the direction of the voluntary innervation. Many examples of apparent movement which can be explained in these terms have been described in the literature. We have selected a few examples for presentation here.

Vestibular Stimulation

In developing our hypothesis concerning the effects of involuntary eye movements on retinal local signs, we have already referred to vestibular stimulation. We discussed a situation in which a stationary object stimulates successive retinal points of the eye which is turning as a result of involuntary innervation of the eye turners. In what follows we are concerned with a situation in which, in spite of vestibular stimulation, the eyes remain stationary.

As was already described, vestibular nystagmus of the eyes can be produced by rotating subjects around their vertical axes. With counterclockwise acceleration the eyes will move slowly clockwise (slow phase of nystagmus). If, under these conditions, an observer wants to keep fixating an object placed in front of him and rotating with him, he must counteract the increasing reflex innervation of his dextrorotators by increasing the voluntary innervation of the levorotators. According to our first hypothesis and its elaboration above, we would predict that such a progressive increase in voluntary innervation of the levorotators would lead to a gradual change in the space value of the fovea. The fixated object would appear to be moving to the left. Morant (18) has reported experiments which bear out this prediction.

Paralysis of the Eye Turners

The clinical literature describes many cases of peripheral paralyses of the eye turners. Patients with such paralyses experience apparent movement of the stimulus field if they try to turn the paralyzed eye. Helmholtz (9) describes a case of a patient whose dextrorotator of the right eye was paralyzed. Observations were made with the paralyzed eye only. The patient fixated an object C (Fig. 3). If he tried to turn his eye to the right, voluntary impulses sent to the dextrorotator did not result in eye rotation. The patient kept fixating C. With the attempted innervation of the dextrorotator the whole stimulus field appeared to the patient to move to the right. Mach (17) has performed an experiment in which he "paralyzed" the eye by mechanical means. He placed putty around the eye in such a way as to immobilize it. If he attempted to turn the eye to the right, he could not do so. Stimuli projected on his retina continued to stimulate the same retinal points. The voluntary innervation of the dextrorotators, however, changed the space values of all retinal points, and the environment appeared to move to the right.

The apparent movement of the environment observed in patients with eye-muscle paralysis and in Mach's experiment occurs in the direction of the intended eye movement. These phenomena were to be predicted from our hypothesis concerning the effects of voluntary innervation of eye turners on retinal local signs.

Autokinetic Movement

Few phenomena have received as much attention from psychologists as the apparent movement of a so-called autokinetic light. As we will deal elsewhere extensively with this special case of apparent egocentric movement, we limit our present treatment of the subject to one short comment.

The erratic course of the stationary autokinetic light has often been ascribed to involuntary eye movements. This theory had to be abandoned when Guilford and Dallenbach (8) photographed the eye movements occurring during the phenomenon and found them to be slight and not of such a nature as to account for the experience. This finding was interpreted subsequently to mean that oculomotor processes have nothing to do with the autokinetic phenomenon (1). Such a conclusion does not appear to be warranted by the evidence. On the contrary, we have just discussed phenomena in which the absence of observable eye movements in the presence of voluntary innervation of eye turners was an explicit condition for apparent movement. We believe that the apparent movement of the autokinetic light belongs in this same class of phenomena.

Induced Movement

If an observer fixates a light, C (Fig. 3*b*), placed straight ahead of him, and if another light, D, is placed to the right of C, the dextrorotators of the eyes receive reflex innervation. This innervation increases as light D is slowly moved to the right. Fixation of C can be maintained only by an increase in the voluntary innervation of the levorotators. According to our first hypothesis, such gradual increase in voluntary innervation of the levorotators leads to a progressive change in the space value of the fovea, and it can be predicted that the fixated object will appear to move to the left. Experimental studies (6) bear out this prediction.

The apparent movement of C is induced by the objective movement of D. On the other hand, if C were moved slowly to the left while a peripheral point, D, was held stationary, apparent movement would be induced in D. As fixation of a slowly moving object is maintained by reflex action (pursuit movement), the space values of the retina would not change (Hypothesis 2). As the eye moves and D remains stationary, D would stimulate more and more peripheral points on the left half of the retina (Fig. 2), thereby appearing to move to the right.

Many instances of such induced movement are known to everyone. The most frequently mentioned example is the apparent movement of the moon seen through drifting clouds. Another example is the familiar illusion experienced when one is sitting in a train and the train on the next track slowly starts to move.

SUMMARY

The relationship between phenomena of visual egocentric localization and processes of involuntary and voluntary oculomotor innervation has been pointed out by many earlier investigators (4, 9, 12, 13, 17, 19). This paper has attempted to describe the nature of this relationship more precisely. From an examination of empirical data relating to visual egocentric localization, two hypotheses were derived to account for the perception of direction and for some phenomena of egocentric apparent movement. Hypothesis 1 states that when dextro- or levorotators of the eye receive voluntary innervation, the space values for points on the retina change in relation to the direction and degree of innervation. Corresponding changes of retinal space values occur with voluntary innervation of the elevators and depressors of the eye. Hypothesis 2 states that when eye position is determined by reflex innervation alone, retinal points do not change their space values. Factors responsible for voluntary and reflex innervation of eye turners were examined, permitting the knowledge of patterns of innervation under various stimulus conditions. Predictions from the hypotheses concerning the effects of voluntary and reflex innervation on space perception were compared with empirical data, and were found tenable for a variety of perceptual experiences.

REFERENCES

1. BARTLEY, S. H. *Vision, a study of its basis.* New York: Van Nostrand, 1941.
2. BEST, F. Die Bedeutung der Hemianopsie fuer die Untersuchung des optischen Raumsinnes. *Arch. f. d. ges. Physiol.,* 1910, **136**, 248–262.
3. BRUELL, J. H. Visual egocentric localization, an experimental study. Unpublished doctor's dissertation, Clark Univer., 1953.
4. CARR, H. A. *An introduction to space perception.* New York: Longmans, Green, 1935.
5. DIETZEL, H. Untersuchungen ueber die optische Lokalisation der Mediane. *Z. f. Biol.,* 1924, **80**, 289–316.
6. DUNCKER, K. Ueber induzierte Bewegung. *Psychol. Forsch.,* 1929, **12**, 180–259.
7. FUCHS, W. Untersuchungen ueber das Sehen der Hemianoptiker und Hemiamblyopiker. *Z. f. Psychol., 2. Abteilung,* 1920, **84**, 76–169.
8. GUILFORD, J. P., & DALLENBACH, K. M. A study of the autokinetic sensation. *Amer. J. Psychol.,* 1928, **40**, 83–91.
9. HELMHOLTZ, H. *Handbuch der Physiologischen Optik,* dritte Auflage, ergaenzt und herausgegeben in Gemeinschaft mit Prof. Dr. A. Cullstrand, Prof. Dr. J.

von Kries, Prof. Dr. W. Nagel. Leipzig: Leopold Voss, 1910.

10. JAFFE, K. M. Effect of asymmetrical position and directional dynamics of configurations on the visual perception of the horizon. Unpublished master's thesis, Clark Univer., 1952.
11. KADEN, S. E. Effect of directional dynamics of configurations and visually perceived words on the position of the apparent horizon. Unpublished master's thesis, Clark Univer., 1953.
12. KLEINT, H. Versuche ueber die Wahrnehmung. *Z. f. Psychol.*, 1937, **141**, 9–44.
13. KLEINT, H. Versuche ueber die Wahrnehmung. *Z. f. Psychol.*, 1938, **142**, 259–290.
14. KOFFKA, K. *Principles of gestalt psychology*. New York: Harcourt, Brace, 1935.
15. LINKSZ, A. *Physiology of the eye*. Vol. 2: *Vision*. New York: Grune & Stratton, 1952.
16. MACDOUGALL, R. The subjective horizon. *Psychol. Rev. Monogr. Suppl.*, 1903, **4**. Pp. 145.
17. MACH, E. *The analysis of sensation*. Chicago: Open Court, 1914.
18. MORANT, R. B. Factors influencing the apparent median plane under conditions of labyrinthian stimulation. Unpublished doctor's dissertation, Clark Univer., 1952.
19. ROELOFS, C. O. Optische Lokalisation. *Arch. Augenheilk*, 1935, **109**, 395–415.
20. SCOBEE, R. G. *The oculorotary muscles*. St. Louis: C. V. Mosby, 1952.
21. WAPNER, S., WERNER, H., BRUELL, J. H., & GOLDSTEIN, A. G. Experiments on sensory-tonic field theory of perception: VII. Effect of asymmetrical extent and starting positions of figures on the visual apparent median plane. *J. exp. Psychol.*, 1953, **46**, 300–307.
22. WERNER, H., & WAPNER, S. Toward a general theory of perception. *Psychol. Rev.*, 1952, **59**, 324–338.
23. WOLTER, J. R. Morphology of the sensory nerve apparatus in striated muscles of the human eye. *Arch. ophthalmol.*, 1955, **53**, 2, 201–207.

(Received for early publication April 1, 1955)

✓

BUREAU OF EDNL. & PSYL. RESEARCH. : :
LIBRARY
25/7/64